ECOLE des HAUTES ETUDES COMMERCIALES

THESE

pour l'obtention du titre de
Docteur es sciences de gestion

Nouveau doctorat (arrêtés du 30 mars 1992, du 21 juillet 1999 et du 25 avril 2002)

SUJET

LES SYSTÈMES MULTI-AGENTS EN MARKETING :

MODÉLISATION PAR LES RÉSEAUX DE PETRI

CANDIDATE

CHRISTINE BALAGUE

JURY

Président

PIERRE-LOUIS LIONS
Professeur au Collège de France

Directeur de thèse

GILLES LAURENT
Professeur, Ecole des Hautes Etudes Commerciales

Rapporteurs

PHILIPPE AURIER
Professeur, Université de Montpellier II

JEAN-LOUIS CHANDON
Professeur, I.A.E. d'Aix-en-Provence
Université Paul-Cézanne-Aix-Marseille III

Suffragants

JANGHYUK LEE
Professeur assistant, Ecole des Hautes Etudes Commerciales

PIERRE-LOUIS LIONS
Professeur au Collège de France

THIBAUT MUNIER
Directeur Général, Millemercis

2005

INTRODUCTION :

L'objectif de cette thèse est l'étude et la modélisation en marketing des systèmes multi-agents qui se développent dans le cadre des nouvelles technologies et en particulier sur Internet. Plus précisément, nous introduisons la méthodologie des réseaux de Petri pour modéliser des systèmes multi-agents interactifs sur Internet.

Dans une première partie, nous étudions le concept de systèmes multi-agents interactifs sur Internet en présentant une revue de littérature s'inspirant de courants de recherche divers. Nous détaillons tout d'abord les recherches en marketing portant sur le concept d'interactivité, en particulier sur sa définition et sur les éléments qu'il regroupe. Nous développons ensuite un deuxième courant de recherche issu de la psychologie sociale qui étudie les interactions des internautes au sein des communautés virtuelles. Le troisième courant théorique que nous présentons provient du domaine de la stratégie et aborde l'interactivité comme source de co-production de valeur pour les entreprises et les consommateurs. Nous détaillons ensuite un courant de recherche très développé, issu de l'intelligence artificielle et de l'informatique, portant sur les systèmes multi-agents. Ce courant définit précisément le concept de systèmes multi-agents. Enfin, nous décrivons les modèles développés en marketing pour étudier le comportement des internautes.

Dans la deuxième partie de cette thèse, nous présentons la méthodologie des réseaux de Petri, modèles développés en recherche opérationnelle et en informatique. Après une introduction de la théorie liée à cette approche, nous présentons sur des cas simples (modèle de base d'une liste de cadeaux, processus de commande d'un client) l'application des réseaux de Petri. Nous terminons ce chapitre par une revue de littérature sur les réseaux de Petri synthétisant les articles utilisant ces modèles sur des problématiques liées au domaine de la gestion.

Les chapitres trois à six de la thèse sont consacrés à des cas d'application. Dans les chapitres trois et quatre, nous appliquons les réseaux de Petri pour modéliser un système de wish list sur Internet. Le chapitre trois détaille notamment les différents éléments du modèle, en particulier chaque bloc de modèle utilisé pour décrire chacun des éléments du système. Le chapitre quatre porte sur la performance du modèle global et sur son extension avec en particulier l'étude de l'impact de campagnes d'e-mailing. Plusieurs simulations de stratégies marketing différentes sont menées dans le cadre de ce chapitre.

Dans les chapitres cinq et six, nous modélisons un jeu promotionnel de marketing viral sur Internet. Le chapitre cinq détaille chacun des blocs de modèles utilisés pour décrire les

différentes étapes du processus. Le chapitre six porte sur la mesure de la performance du modèle global et sur plusieurs simulations menées afin de déterminer l'impact de différentes stratégies marketing sur la performance du jeu en terme d'efficacité de marketing viral. Après ces deux applications, nous concluons la thèse en détaillant les avantages des réseaux de Petri, les limites du travail effectué et l'ensemble des pistes de recherche futures ouvertes par l'application potentielle des réseaux de Petri au domaine du marketing.

Au jury de thèse :

A Gilles Laurent,
Professeur au groupe HEC,
directeur de recherche :

Nous sommes fière d'avoir été votre élève tout au long de cette thèse que vous avez imprégnée de vos conseils, de votre grande compétence académique, votre constante rigueur, votre disponibilité et vos profondes qualités humaines.
Que ce travail soit le témoignage de notre profond respect et de toute notre admiration.

A Jean-Louis Chandon,
Professeur, IAE d'Aix-en-Provence,
Université Paul-Cézanne-Aix-Marseille III,
rapporteur,

A Philippe Aurier,
Professeur à l'Université Montpellier II,
rapporteur:

Qui nous font l'honneur de participer à notre jury de thèse.
Nous vous remercions particulièrement d'avoir accepté d'être rapporteurs et de bien vouloir juger notre travail.
Veuillez croire à l'expression de notre reconnaissance et de notre respect.

A Pierre-Louis Lions, Professeur au Collège de France:

En hommage respectueux pour l'honneur que vous nous faites de siéger à la soutenance de notre thèse.

A Janghyuk Lee, Professeur assistant au groupe HEC :

Qui s'est intéressé à nos travaux et nous a initiée à la publication de nos recherches.
Pour les conseils et pour son soutien qu'il nous a toujours accordés avec gentillesse et amitié.
Sincères remerciements.

A Thibaut Munier, Directeur Général, MilleMercis :

Qui nous a cordialement fourni des données particulièrement intéressantes pour l'application de nos modèles.
Pour sa collaboration dans le cadre de nos recherches,
Pour son soutien et ses remarques lors des étapes intermédiaires,
Que ce travail soit le témoignage de notre sincère reconnaissance.

A Marc Vanhuele,

Pour la richesse de ses idées, la rigueur de ses analyses, sa passion de la recherche qu'il nous a communiquées.
Qu'il trouve ici le témoignage de notre gratitude.

A Jean-Michel Lasry,

Qui s'est intéressé à nos travaux et nous a constamment soutenu de son amitié.
Pour les conseils qu'il nous a toujours accordés avec gentillesse,
Pour sa passion de la recherche,
Qu'il trouve ici le témoignage de notre respectueuse admiration et affection.

A Jean-Michel, Hadrien, Eléna, Matéo et Léonore,

Pour leur constante et affectueuse pression à comprendre le monde qui nous entoure et celui de demain.

A mes Parents,

Pour leur soutien permanent depuis toujours.
Avec toute ma tendresse.

A tous nos amis du doctorat,
A Marie-Dominique de Saint-Aubert, Marie-Anne Dureuil, Elisabeth Sartiaux,

Pour leur soutien et pour les bons moments passés ensemble.
Amicales pensées.

SOMMAIRE

Introduction p. 1

Chapitre 1: les systèmes interactifs multi-agents sur Internet : revue de littérature p. 3

1.1.1 Introduction p. 3
1.1.2 Les recherches en marketing sur l'interactivité p. 4
1.1.3 L'interactivité: un concept encore peu précis dans le domaine professionnel p. 4
1.1.4 Les théories en Marketing sur l'interactivité p. 5
1.1.4.1 L'interactivité du point de vue de l'entreprise p. 8
1.1.4.2 L'interactivité du point de vue du consommateur p. 12
1.1.4.3 Les facteurs constitutifs de l'interactivité p. 16
1.1.4.4 L'interactivité comme processus p. 24
1.1.5 L'interactivité entre internautes analysée par des approches en psychologie sociale p. 25
1.1.6 Les groupes sociaux en psychologie sociale p. 26
1.1.7 Le marketing tribal p. 31
1.1.8 Les communautés virtuelles p. 35
1.1.9 L'interactivité en stratégie comme source de création de valeur p. 49
1.1.10 Les processus interactifs dans les systèmes multi-agents p. 52
1.1.11 Les modèles de comportement des internautes dans la littérature en marketing p. 63
1.1.12 Les recherches sur l'amélioration des sites Web p. 63
1.1.13 Les modèles de comportement de visite et d'achat sur le Web p. 66
Conclusion p. 74

Chapitre 2 : les réseaux de Petri : modèle et revue de littérature p. 76

2.1 Introduction aux réseaux de Petri p. 76
2.2 Présentation des réseaux de Petri p. 77
2.2.1 Eléments de base p. 77
2.2.2 Exemple : modélisation d'un mécanisme de wish list sur Internet (modèle de base) p. 78

2.2.3 Deuxième exemple: description du processus de commande d'un client p. 87
2.2.4 Les différentes familles de réseaux de Petri p. 99
2.3 Les réseaux de Petri stochastiques p.101
2.3.1 Présentation p.101
2.3.2 Eléments de base d'un SPN et exemple p.104
2.4 Littérature sur les réseaux de Petri p.110
2.4.1 Meldman (1977) : première introduction des réseaux de Petri dans le domaine du management p.111
2.4.2 Les réseaux de Petri utilisés pour modéliser des systèmes de flux p.113
2.4.3 Les réseaux de Petri pour modéliser des systèmes d'organisation ou des systèmes multi-agents : p.119
Conclusion p.123

Chapitre 3 : Application des Réseaux de Petri à la modélisation des systèmes de wish list p.124

3.1 Introduction p.124
3.2 Cadre conceptuel p.124
3.3.1 Cadre méthodologique des réseaux de Petri pour modéliser un système de wish list p.127
3.3.2 Le concept de wish list p.127
3.3.3 Les données utilisées p.129
3.3.4 Modèle de Petri appliqué aux wish lists p.130
3.4 Détail du modèle p.132
3.4.1 Loi de durée interconnexion des agents A p.133
3.4.2 Loi de durée interconnexion des agents B p.167
3.4.3 Loi d'achat des agents A ou B p.171
Conclusion p.188

Chapitre 4 : Les réseaux de Petri pour modéliser les systèmes de wish list : performance du modèle et extension à la mesure de l'impact de l'e-mailing p.189

4.1 Introduction p.189
4.2 Résultats des réseaux de Petri pour modéliser les wish lists p.189
4.2.1 Sélection des données pour le calcul des paramètres des lois stochastiques p.190
4.2.2 Performance du modèle du réseau de Petri p.214
4.2.3 Comparaison du modèle du réseau de Petri par rapport à un modèle NBD p.218
4.3 Extension à la mesure de l'impact de l'e-maling et implications managériales p.227
4.3.1 Intérêt de modéliser l'e-mailing p.229
4.3.2 Modélisation de l'e-mailing par les réseaux de Petri p.230
4.3.3 Modèle HLM pour déterminer le taux d'ouverture d'un e-mail p.231
4.3.4 Intégration de l'impact de l'e-mail dans le réseau de Petri p.241
4.3.5 Implications managériales p.243
Conclusion p.250

Chapitre 5 : Modélisation d'un jeu viral promotionnel on line par les réseaux de Petri p.253

5.1 Introduction p.253
5.2 Revue de littérature sur le marketing viral sur le Web p.254
5.3 Principe de l'opération « Le ciel est à tout le monde » p.260
5.4 Modélisation du jeu promotionnel « Le ciel est à tout le monde » par les réseaux de Petri p.266
5.4.1 Modèle du nombre d'inscriptions de joueurs par jour p.269
5.4.1.1 Modélisation du nombre d'inscriptions venant de l'affiliation p.274
5.4.1.2 Modélisation du nombre d'inscriptions venant du mail d'annonce p.290
5.4.1.3 Modélisation du nombre d'inscriptions venant du mail de relance p.303
5.4.1.4 Modélisation du nombre d'inscriptions venant de l'autopromotion p.312
5.4.1.5 Modélisation du nombre d'inscriptions venant de source non identifiée p.316

5.4.2 Modèle des non-réponses aux questions p.322
5.4.2.1 Modèle continu de non-réponse aux questions p.325
5.4.2.2 Modèle discret de non-réponse aux questions p.353
5.4.3 Modèle de la transformation d'adresses virales saisies en adresses virales joueurs p.367
5.4.3.1 Loi de la probabilité qu'une adresse virale saisie joue p.368
5.4.3.2 Durée entre saisie adresse virale et inscription au jeu de l'adresse virale p.375
Conclusion p.377

Chapitre 6: Modélisation, mesure de la performance et simulation d'un jeu promotionnel viral on line par les réseaux de Pétri p.378

6.1 Introduction p.378
6.2 Réseau de Petri du jeu promotionnel on line p.378
6.2.1 Synthèse des modèles à intégrer dans le réseau de Pétri p.381
6.2.2 Performance du modèle p.382
6.2.2.1 Performance du modèle sur la partie inscription des joueurs p.383
6.2.2.2 Performance globale du modèle p.387
6.3 Simulations p.390
6.3.1 Simulations sur les inscriptions sur 3 périodes de jeu (10, 20 et 30 jours) p.392
6.3.1.1 Simulation sur l'affiliation p.392
6.3.1.2 Simulation sur le mail de relance p.395
6.3.1.3 Simulation sur l'autopromotion p.397
6.3.1.4 Simulation sur un mix de techniques de recrutement p.400
6.3.2 Simulations sur le design du questionnaire p.402
6.3.2.1 Simulation 10 questions alternant 1 question de qualification / 1 question virale p.402
6.3.2.2 Simulation 10 questions alternant 2 questions de qualification/ 2 virales p.404
6.3.2.3 Simulation 10 questions avec 5 questions de qualification suivies de 5 questions virales p.407
6.3.2.4 Simulation 10 questions avec 5 questions virales suivies de 5 questions de qualification p.409
6.3.2.5 Synthèse des simulations testées et implications managériales p.411
Conclusion p.414

Conclusion finale **p.415**

Bibliographie **p.431**

Annexe 1 : Analyse comportementale des réseaux de Petri **p.443**

Annexe 2 : PP plots durée interconnexion agents A **p.452**

Annexe 3 : Loi de la durée interconnexion agent A : cas particuliers **p.487**

Annexe 4 : Loi de la durée interconnexion agent B **p.497**

Chapitre 1:

Les systèmes interactifs multi-agents sur Internet : revue de littérature

1.2 INTRODUCTION :

Les systèmes interactifs multi-agents sur Internet constituent un thème de recherche dans des domaines aussi divers que le marketing, la psychologie sociale, la stratégie, l'intelligence artificielle, ou encore l'informatique et la recherche opérationnelle. L'objectif de ces recherches est d'étudier les nouveaux environnements liés au développement de l'Internet et plus particulièrement les interactions entre les différents agents qui agissent au sein de ces environnements. Ce chapitre présente une revue de littérature issue de différents domaines permettant de mieux définir et comprendre les systèmes interactifs multi-agents sur Internet.
En effet, chaque courant de recherche que nous allons présenter dans cette revue de littérature va nous éclairer sur l'un ou l'autre des termes utilisés dans la notion de systèmes interactifs multi-agents sur Internet et va nous permettre d'affiner par conséquent la définition de ce concept. Nous allons dans une première partie décrire la littérature en marketing concernant le concept d'interactivité et ses mesures. Dans une deuxième partie, nous développerons les approches en psychologie sociale étudiant en particulier certains systèmes interactifs multi-agents sur Internet, en particulier les interactions entre internautes au sein de communautés virtuelles. La troisième partie sera consacrée aux théories développées en stratégie sur la co-production de valeur créée par l'interaction entre plusieurs agents ou entités. Le quatrième champ synthétise les nombreuses recherches en informatique et en intelligence artificielle sur les systèmes multi-agents et présente les théories développées dans ces domaines pour étudier les interactions entre plusieurs entités d'un même environnement. Enfin, dans la dernière partie, nous présenterons un état des lieux de l'ensemble des modèles utilisés en marketing pour mesurer et analyser le comportement des internautes sur Internet, en particulier les modèles de comportement de visite et d'achat sur une site Web, qui permet de modéliser les interactions entre l'internaute et Internet. Nous conclurons ensuite sur les limites actuelles des recherches issues de la littérature exposée dans ce chapitre par rapport à l'analyse des systèmes multi-agents sur Internet dans le domaine du Marketing, et sur la nécessité d'introduire de nouvelles approches pour modéliser de tels systèmes.

1.3 LES RECHERCHES EN MARKETING SUR L'INTERACTIVITE:

1.3.1 L'interactivité : un concept encore peu précis dans le domaine professionnel

Le développement rapide de l'Internet à la fois chez le consommateur et dans les entreprises a ces dernières années modifié la communication des entreprises envers leur cible et les stratégies marketing. Historiquement, cette évolution des entreprises s'est effectuée au cours de deux périodes distinctes. La première correspond au développement de la bulle Internet et de l'emballement high-tech du milieu des années 90. Le monde du e-commerce était alors le centre de toutes les attentions, au centre des préoccupations des media, des marchés financiers, du marketing, des managers en communication et même du top management. Face à cet engouement et au potentiel de l'Internet , on aurait pu penser que le Web allait être au centre des stratégies des managers. En réalité, la plupart des entreprises ont durant cette période plutôt testé Internet. La démarche courante a été de créer un site Web afin d'avoir un minimum de présence virtuelle, par crainte de rester enfermé dans l'ancienne économie (Rayport & Jaworski, 2002, p.2). Les activités Internet ont pourtant été à cette période des sources de pertes financières pour les entreprises. A part un nombre limité de business models à succès (e-bay, amazon), le e-commerce durant ces années a été davantage l'objet d'un grand battage publicitaire que d'un travail de fond. L'explosion de la bulle Internet d'Avril 2000 a montré la vulnérabilité des premiers explorateurs de ce nouveau monde. La deuxième période concerne les années post avril 2000, encore en cours. Internet a continué de se développer de manière forte auprès des consommateurs et des entreprises, comme une vraie révolution de fond toujours en marche. En revanche, les stratégies se sont recentrées de manière rationnelle sur le vrai rôle de l'Internet et de l'e-commerce. Un travail de fond a été mené par la plupart des entreprises et le Web est devenu aujourd'hui un canal souvent incontournable pour une entreprise. L'une des raisons est l'interactivité que permet le nouvel environnement du Web et qui lui est spécifique. Pourquoi donc le concept d'interactivité a-t-il résisté à l'effondrement de la bulle Internet ? La réponse se trouve probablement dans le fait que l'interactivité est au cœur de la réelle révolution apportée par Internet, entraînant une profonde évolution des outils marketing. Aujourd'hui, le mot "interactivité" est souvent utilisé par les chercheurs, les responsables marketing, les agences de site Web, les consommateurs et les media. Se créent actuellement des agences interactives ou encore des agences de marketing interactif. On assiste même dans les entreprises à l'apparition d'une nouvelle fonction : directeur de l'interactivité ou directeur des systèmes interactifs. Le cabinet Bates/Alexander & Associates

a même développé un Index Telemedia Interactif (ITI) pour mesurer l'interactivité de l'environnement media (TV ou Internet) dans plusieurs secteurs d'activité (fournisseurs de contenu, commerce électronique, équipement informatique, transports, secteur public) (Gugel, 1997). Cependant, chez les praticiens en marketing, la signification concrète du mot interactivité reste peu claire. On constate qu'il désigne des éléments très divers, plus ou moins précis : les agences dites interactives sont généralement des agences dont l'activité est de créer des sites Web (sans forcément utiliser les mécanismes d'interactivité du Web), le directeur de l'interactivité est souvent un directeur e-business, le marketing interactif est souvent limité à un ensemble de techniques cherchant à créer une nouvelle relation interactive avec le consommateur. Il n'y a pas réellement de définition claire et homogène de la notion d'interactivité au sein des managers. Par conséquent, on constate que l'interactivité est un concept peu maîtrisé par les praticiens et que des recherches sont nécessaires pour mieux définir, clarifier et préciser le concept . Nous allons ci-dessous présenter un état des lieux des théories développées en Marketing sur le concept d'interactivité.

1.3.2 Les théories en Marketing sur l'interactivité :

Plusieurs articles dans le domaine du Marketing portent sur l'interactivité, le concept et ses mesures. On constate cependant que ces articles diffèrent les uns des autres dans la manière d'appréhender ce concept et que l'interactivité ne constitue pas un courant de recherche spécifique et homogène.

Certains auteurs (Coviello, Milley et Marcolin, 2001) analysent globalement l'impact de l'interactivité sur le marketing contemporain pratiqué par les entreprises. Ils affirment que l'interactivité a le potentiel d'affecter tous les business et tous les types de produits, avec des techniques modifiant profondément les liens entre les entreprises et leurs marchés. Selon eux, l'interactivité facilite et simplifie ce lien tout en constituant un nouveau canal de distribution. Ces auteurs classifient l'interactivité en cinq approches marketing: le marketing de la transaction, les bases de données, le marketing de l'interaction, le marketing des réseaux, et le e-marketing. Chaque approche possède ses spécificités et Coviello, Milley et Marcolin (2001) soulignent qu'une entreprise peut utiliser ces techniques face à différents consommateurs dans un processus d'échange. Selon eux, le e-marketing est parmi ces techniques celle qui a la plus forte capacité de transformer un processus d'interactivité technique en business rentable.

McMillan et Hwang (2002) effectuent une revue de la littérature approfondie de toutes les définitions de l'interactivité et les regroupent en quatre familles: les définitions décrivant le processus d'interactivité, celles présentant les caractéristiques de l'interactivité, les définitions portant sur la perception de l'interactivité et enfin les articles combinant processus, caractéristiques et perception de l'interactivité. Ils montrent ainsi la multitude des approches proposées et en tirent trois éléments apparaissant fréquemment sur l'interactivité: la bi-direction de la communication, le contrôle de l'utilisateur, le temps. Notre analyse de la littérature, s'appuyant sur ces recherches mais intégrant d'autres approches non incluses dans McMillan et Hwang (2002), nous permet d'identifier quatre courants de recherche sur l'interactivité que nous avons synthétisés dans le tableau 1.1 ci-dessous : l'interactivité du point de vue de l'entreprise ; l'interactivité du point de vue du consommateur ; les facteurs constitutifs de l'interactivité ; le processus d'interactivité. La première conclusion issue du tableau 1.1 est que l'interactivité en Marketing est un concept multidimensionnel.

Tableau 1.1 : Les principaux courants de recherche sur l'interactivité en Marketing

Thèmes	Auteurs principaux	Concept de l'interactivité
Interactivité du point de vue de l'entreprise - Database - Communication - Site Web	Blattberg et Deighton (1991) Deighton (1996) Ghose et Dou (1998)	Database, « Adressabilité » Communications à double sens : -capacité de s'adresser à un individu -capacité à réunir, à se souvenir des réponses de cet individu et à lui fournir une solution adaptée à ces réponses, critères d'interactivité d'un site
Interactivité du point de vue du consommateur	Card, Moran and Newell (1983) Steuer (1992)	Modèle d'interaction homme-ordinateur -Degré auquel les utilisateurs d'un medium peuvent influencer la forme ou le contenu de son environnement

		-Téléprésence
	Reeves et Nass (1996))	Psychologie sociale
Facteurs constitutifs de l'interactivité	Steuer (1992)	3 facteurs : -Vitesse -Etendue -Mapping
	Cho and Leckenby (1997)	interaction user-machine interaction user-user interaction user-message
	Burgoon, Bonito, Bengtsson et Ramirez (1999/2000)	Interactivité à partir d'éléments structurels ou par des expériences qualitatives
	Yuping Liu and LJ Shrum (2001)	3 dimensions : Active control 2 way communication Felt synchronicity
Interactivité comme processus	Csikszentmihalyi (1990)	Processus d'expérience optimale : attitude du consommateur utilisant un environnement CME
	Beszian-Avery, Calder et Iacobucci (1998)	Processus iteratif par lequel les désirs et besoins du consommateur sont couverts, puis modifiés et satisfaits par le producteur Arbre hiérarchique / circulaire/ réseau

Nous allons détailler ci-dessous ces quatre grands courants de recherche, avant de présenter une synthèse des méthodologies utilisées dans la littérature pour mesurer l'interactivité sur le Web.

1.3.2.1 L'interactivité du point de vue de l'entreprise :

Les premiers articles en marketing sur l'interactivité datent du début des années 90, période pendant laquelle deux auteurs en particulier (Blattberg et Deighton (1991)) ont commencé à publier sur ce concept. Blattberg et Deighton (1991) analysent le marketing interactif comme la capacité à développer des relations interactives grâce à l' « adressabilité » des consommateurs individuels permettant de construire des bases de données et d'entretenir le dialogue.

L'interactivité devient dans les recherches postérieures un élément de communication pour l'entreprise: Deighton (1996) présente l'interactivité comme ayant deux types de communication : la capacité de s'adresser à un individu plus la capacité à réunir et à se souvenir des réponses de cet individu, ces deux caractéristiques rendant possible une troisième: la capacité de s'adresser à nouveau à l'individu en tenant compte de la réponse de celui-ci. Deighton (1996) affirme ainsi que l'interactivité est un outil permettant de transformer un bon marketing en une bonne conversation, lui donnant ainsi un visage humain.

Plusieurs autres chercheurs ont montré le rôle de l'interactivité dans les efforts des entreprises pour construire de bonnes relations avec leurs consommateurs. Hoffman, Novak et Chatterjee (1995) soulignent le fait que le Web libère les consommateurs de leur rôle traditionnellement passif de récepteurs de communications marketing et leur donne un plus grand contrôle sur la recherche d'information et le processus d'acquisition en leur permettant d'être des participants actifs du processus marketing. Upshaw (1995) affirme que la nature interactive du Web offre aux marketers de nouvelles opportunités pour créer de plus fortes identités de marque ayant le potentiel de développer la fidélité à la marque. Cuneo (1995) développe l'idée selon laquelle le potentiel pour interagir avec les consommateurs facilite le marketing relationnel et les supports aux consommateurs à un degré jamais égalé par d'autres media traditionnels (TV, radios, etc…). Berthon et al. (1996) suggèrent que le niveau d'interactivité d'un site est un élément essentiel pour la conversion de visiteurs intéressés du site en consommateurs interactifs.

Ghose & Dou (1998) montrent quant à eux que plus le degré d'interactivité d'un site est important, plus il appartient à la liste renommée des « top » sites effectuée par Lycos (liste Lycos Top 5%). Ils utilisent pour mesurer le concept d'interactivité les outils interactifs employés dans la nomenclature IPS (Internet Presence Site). L'IPS est communément utilisée (Hoffman et al. 1995) pour définir les éléments d'un site Web Corporate fournissant une présence virtuelle de l'entreprise et présentant les différentes offres de l'entreprise. Au sein de cette nomenclature, Ghose & Dou (1998) sélectionnent 23 outils interactifs utilisés en marketing communication et les regroupent autour de cinq grandes fonctions marketing: support client, marketing research, aide personnalisée au consommateur, communication/promotion/publicité, jeu/loisir. La liste de ces outils est fournie dans le tableau 1.2 ci-dessous.

Tableau 2.2: Les 23 outils interactifs de la nomenclature IPS (Ghose and Dou, 1998)

Customer Support	Software downloading: surfers download software from a site, usually for free. Online problem diagnostics: customers report their problem spots and this function helps them to locate the problem exactly. Whenever possible, "trouble shooting" suggestions are given. Electronic form (e-form) inquiry: e-forms on which customers can type in online inquiries regarding the products or the firm. Order status tracking: customers can track the status or whereabouts of their orders online in real time. Comment: customers can fill out e-forms to express their opinions about the company, products, and the site. Feedback: customers can type in their feedback in e-forms with regard to specific questions raised by the site.
Marketing Research	Site survey: e-form survey for visitors that solicits their comments on the content and design of the site. Product survey: e-form survey designed for measuring customer satisfaction about firm's offerings and service. New-product proposal: e-forms for customers to write about their expectations of new products and their suggestions for new products.

Personal Choice-Helper	Key word search: a function that allows a visitor to pinpoint the particular information he or she is interested in. Personal-choice helper: a function that can make relatively sophisticated recommendations on consumers' choices based on their input of preferences and decision criteria. Virtual reality display: a function that permits consumers to virtually "feel or experience" the product. Dealer locator: a function that allows users to pinpoint a dealer closest to his or her residence.
Advertising/ Promotion/Publicity	Electronic coupon: distributed online and can be used in retail stores. Usergroups: cyber community for product users. Online order an option to order products online. Sweepstakes/prize: events held to attract surfers and to encourage surfer participation by special incentives. Multimedia shows: quicktime movie streamline video, and other forms of multimedia presentations. Push media: similar to TV channels. Users select to participate and receive information directly to their screens on a regular basis. Interactive job placement: online resume building, personal career goal check, etc…

Entertainment	Electronic post card: written by senders online and to be retrieved by recipients. Surfer postings: a section for surfers to write their stories, opinions, or others. Games: online games.

A partir des outils interactifs listés dans le tableau 1.2, les auteurs construisent un Index d'Interactivité (II) permettant de quantifier le niveau d'interactivité qu'un internaute peut expérimenter sur un site. Utilisant les techniques d'analyse de contenu, Ghose & Dou (1998) analysent ensuite le contenu de 101 sites Web de fabricants de produits de grande consommation et de biens durables. Ils développent ensuite un modèle Logit basé sur l'Index d'Interactivité pour déterminer l'appartenance ou non d'un site à la liste Lycos Top 5%.
Les principaux apports de cette recherche sont les suivants. D'une part, les auteurs synthétisent différentes fonctions interactives utilisables en marketing électronique; d'autre part ils proposent un Index d'Interactivité simple à utiliser et montrent que plus les sites ont un Index d'Interactivité élevé, plus leur probabilité d'appartenir à la liste Lycos Top 5% est forte.

1.3.2.2 L'interactivité du point de vue du consommateur :

On distingue dans ce courant de recherche en marketing deux grands groupes d'auteurs. Les premiers ont analysé le comportement du consommateur et sa réaction face à des processus interactifs, les seconds ont développé des approches expérientielles.
Les premières recherches datent d'une vingtaine d'années : Card, Moran et Newell (1980) s'intéressent aux interactions homme-ordinateur et développent un modèle tenant compte du temps passé et de plusieurs autres facteurs physiques influençant la performance de l'utilisateur.
Le comportement du consommateur dans un environnement interactif est par la suite analysé en marketing par Steuer (1992) qui définit l'interactivité comme le degré auquel les utilisateurs d'un médium peuvent en influencer la forme et le contenu. Steuer dans un premier temps parle de réalité virtuelle en introduisant le concept de présence comme facteur clef de cette réalité virtuelle. Il définit la présence comme la sensation d'être dans un environnement.

Lorsque cette perception est modifiée par une technologie de communication, l'utilisateur est obligé de percevoir simultanément deux environnements: l'environnement physique dans lequel il est réellement présent et l'environnement du médium lui même. Steuer introduit ensuite le concept de télé-présence pour décrire le degré de présence ressenti par l'utilisateur dans un environnement avec médium par rapport à celui ressenti dans un environnement physique réel. Il définit donc la télé-présence comme l'expérience de présence dans un environnement grâce à un médium de communication. En d'autres termes, la présence se réfère à la perception naturelle de l'environnement, tandis que la télé-présence se réfère à la perception médiatisée de l'environnement. Cet environnement peut être un environnement temporellement réel ou spatialement distant (un endroit lointain vu avec une caméra vidéo par exemple) ou un monde virtuel non existant synthétisé par ordinateur. Selon Steuer, ces nouveaux concepts suggèrent une nouvelle vision de la communication. Traditionnellement, un processus de communication se décrit comme la transmission d'information d'un émetteur à un récepteur. La télé présence a contrario se focalise sur la relation entre un individu pouvant être à la fois émetteur et récepteur et l'environnement médiatisé avec lequel il interagit, comme le montre la figure 1.1. L'information n'est donc pas transmise d'un émetteur à un récepteur. On assiste plutôt à la création d'environnements médiatisés dans lesquels les individus vivent des expériences.

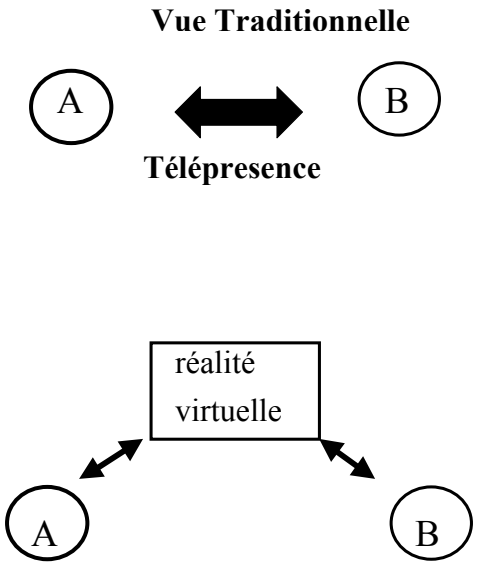


Figure 1.1 :
Deux modèles de communication avec médium intermédiaire
(Steuer, 1992)

Dans un deuxième temps, Steuer définit les deux déterminants de la télé-présence. Le premier est la vivacité qu'il définit comme la capacité d'une technologie à produire un environnement médiatisé sensoriellement riche. Les deux facteurs qui contribuent à la vivacité sont l'ampleur et la profondeur. L'ampleur correspond au nombre de dimensions sensorielles simultanément présentes et la profondeur à la définition de chaque canal de perception. Le second déterminant de la télé-présence est l'interactivité que Steuer définit comme le degré de participation des utilisateurs d'un médium qui peuvent en modifier la forme ou le contenu en temps réel. L'interactivité selon Steuer est donc déterminée par la structure technologique du médium.

Les recherches de Steuer diffèrent de celles d'autres auteurs comme Rafaeli (1988) qui définit l'interactivité dans une vision de communication plus traditionnelle. Selon cet auteur, l'interactivité est une variable caractéristique de la composition d'une communication. L'interactivité exprime le degré d'une série d'échanges de communication, toute transmission supplémentaire étant reliée aux précédentes en fonction du niveau auquel ces précédents échanges étaient eux mêmes reliés à des transmissions encore antérieures.

Les interactions internaute-ordinateur ont été étudiées avec profondeur par Reeves et Nass dans « The Media Equation » (1996), ouvrage du domaine de la psychologie sociale dans lequel les auteurs étudient au travers d'expériences les facteurs explicatifs des réactions des utilisateurs face aux média TV ou Internet. Contrairement à une croyance très répandue, Reeves et Naas montrent que les interactions des individus avec les ordinateurs, la télévision et les nouvelles technologies de communication sont fondamentalement sociales, naturelles, et qu'elles sont identiques aux relations sociales réelles entre les hommes. Grâce à plusieurs études fondées sur des théories en psychologie sociale, les auteurs affirment que les individus traitent les ordinateurs, la télévision et les nouveaux media comme des personnes ou des lieux réels. Ils posent par conséquent l'équation des media : « media=vie réelle ». Reeves et Naas mènent des expériences sur cinq thèmes concernant les interactions homme-ordinateur: media et manières ; media et personnalité; media et émotion; media et rôle social; media et formes. Les expériences menées sur le premier thème media et manières apportent plusieurs conclusions. Tout d'abord, les relations homme-ordinateur sont influencées par la politesse, suggérant ainsi l'importance de développer des communications interactives intégrant ce discours de politesse (dire bonjour, répondre, tenir compte des différences culturelles, etc…). L'interaction homme-ordinateur est également influencée par la distance interpersonnelle

(distance entre la personne et le message) qui détermine la durée de l'interaction et son intensité émotionnelle. Concrètement, ceci se traduit par un impact positif de la taille des écrans, des photos, des distances de vue sur l'attention des individus. La flatterie et le jugement des autres constituent deux facteurs explicatifs d'une forte interaction homme-ordinateur: les utilisateurs aiment être félicités, détestent être critiqués (notamment dans les messages d'erreur souvent fournis par les ordinateurs). Le deuxième thème abordé par Reeves et Naas porte sur la personnalité. Les auteurs montrent que dans la majorité des interactions, la personnalité des intervenants est un élément fondamental. Testant plusieurs personnalités dans les interfaces des ordinateurs (personnalité dominante ou soumise par exemple), ils montrent un impact positif sur l'interaction lorsque la personnalité de l'utilisateur est proche de celle fournie par l'ordinateur. Une implication concrète de cette étude est la création de personnages virtuels ou d'objets sur les sites présentant des traits de personnalité. En ce qui concerne le troisième thème des émotions, Reeves et Naas montrent que les interactions hommes-ordinateurs sont influencées par l'évaluation que les utilisateurs font du contenu du médium en « bien ou mal ». Ils effectuent également des expériences sur la négativité (envoi de messages négatifs) ayant un fort impact sur la mémorisation du contenu des messages. Les auteurs montrent également que la stimulation et l'excitation des émotions influencent l'interaction. Le quatrième thème qu'ils étudient concerne les rôles sociaux. Une première conclusion issue de leurs expériences est que les consommateurs face à un médium sont influencés par les témoignages d'experts ou de spécialistes. Les auteurs s'intéressent ensuite à l'influence de la notion de co-équipiers sur l'interaction homme-ordinateur. Pour cela, ils mènent plusieurs expériences en constituant des équipes d'une personne et d'un ordinateur, chaque équipe possédant une identité d'équipe (nom, couleur, logo). Le principal résultat est que l'interaction homme-ordinateur est plus importante dans le cas d'une équipe homme-ordinateur que dans les situations sans équipe créée. Reeves et Naas analysent également l'impact de voix féminines ou masculines sur l'interaction, montrant que les gens répondent à différentes voix comme à différents acteurs sociaux. Un autre facteur d'influence est l'orientation de la source ou encore l'association entre les sources et le contenu : les personnes assignent automatiquement la responsabilité des messages à ceux qui les délivrent même si le récepteur sait que ce lien est douteux. Enfin, le cinquième thème traité concerne les formes : Reeves et Naas mènent plusieurs expériences sur la taille des images, l'acuité visuelle (images nettes ou non), la synchronisation audio-vidéo, le mouvement, les changements de scènes, les images subliminales et démontrent l'impact de ces facteurs sur les interactions hommes-ordinateur.

1.3.2.3 Les facteurs constitutifs de l'interactivité :

Plusieurs recherches en marketing portent sur les facteurs constitutifs de l'interactivité. On constate cependant qu'il n'existe pas vraiment de consensus homogène et reconnu par la communauté des chercheurs sur les dimensions de l'interactivité. Ceci suggère que de futures pistes de recherches restent ouvertes sur ce thème de recherche. Les principales recherches sur les dimensions de l'interactivité sont décrites ci-dessous.

Après avoir défini la réalité virtuelle et la télé-présence (cf paragraphe 1.2.2.2), Steuer (1992) détermine trois facteurs influençant l'interactivité d'un médium: la vitesse, définie comme la rapidité avec laquelle une entrée de données est assimilée par l'environnement du médium; le champ, défini comme le nombre de possibilités d'une action à tout moment ; la cartographie, définie comme la capacité du système à dresser naturellement et de manière prévisible un plan de contrôle des changements au sein de l'environnement média. Ces éléments sont synthétisés dans la figure 1.2 ci-dessous.

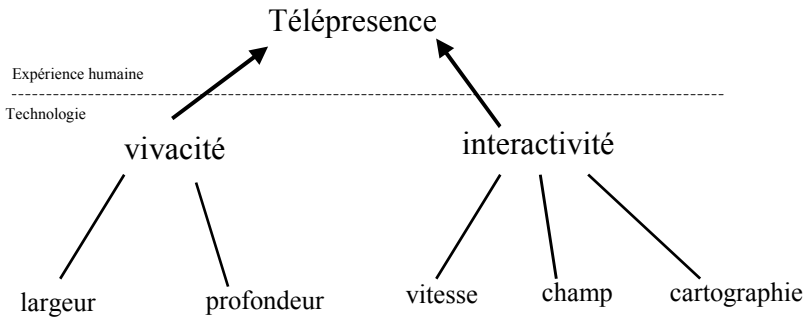


Figure 1.2: Variables technologiques influençant la téléprésence
(adapté de Steuer, 1992)

Selon Steuer, la vitesse d'interaction est une caractéristique importante d'un médium interactif, l'interaction en temps réel représentant la plus forte valeur possible de cette variable. Plusieurs nouveaux media essaient d'atteindre ce niveau d'interactivité, avec dans

certains secteurs comme les jeux vidéo des performances techniques élevées visant à ce que l'utilisateur ait la perception du monde en temps réel. Le champ de l'interactivité décrit par Steuer se rapporte à l'ensemble des modifications possibles du médium à tout moment par l'utilisateur. Les dimensions spécifiques qui peuvent être modifiées dépendent des caractéristiques du médium, incluant la possibilité de stopper une image ou de continuer le processus (vidéo, défilement d'images) à tout moment, l'organisation spatiale (objets qui apparaissent), l'intensité (sons, clarté des images), les couleurs. Plus ces paramètres peuvent être modifiés, plus l'interactivité est forte. Steuer souligne le fait que plusieurs recherches ont été menées sur l'interactivité dans le domaine de l'interaction homme-machine, mais que peu de recherches ont été menées sur ce thème dans le domaine de la communication.

Dans le courant de recherche de Steuer, d'autres auteurs ont mené des études empiriques. Coyle et Thorson (2001) analysent l'impact de niveaux progressifs d'interactivité sur l'attitude des individus vis-à vis du site. Ils montrent d'une part que, plus l'interactivité et la vivacité (selon la définition de Steuer, 1992) sont importantes, plus la sensation de téléprésence est forte. D'autre part, une plus grande vivacité est associée à des attitudes plus favorables vis-à-vis du site Web.

Deux autres auteurs, Ha et James (1998), utilisent l'analyse de contenu pour étudier le niveau d'interactivité de plusieurs sites Web. Ils détectent cinq dimensions de l'interactivité : l'enjouement (le fait de pouvoir jouer avec le site), l'interconnexion, la collecte d'information, la réciprocité de la communication et les choix proposés.

Cho et Leckenby (1997) proposent une synthèse et une classification des définitions de l'interactivité fournies par la littérature. Ils regroupent les recherches en trois thèmes: les interactions user-machine (ce sont historiquement les premières et les plus fréquentes approches de l'interactivité traitées dans les recherches), les interactions user-user (communications interpersonnelles) et les interactions user-message (capacité de l'utilisateur à contrôler et à modifier le message, Steuer, 1992).

Raman et Leckenby (1998) dans une recherche ultérieure montrent que l'interaction sur Internet a un impact positif sur la durée de visite du site. Pour mesurer l'interaction sur Internet, ils développent une échelle fondée sur d'autres échelles existantes d'utilisation de l'ordinateur. Cette échelle est présentée dans le tableau 1.3 ci-dessous. On constate notamment que les items de cette échelle ne correspondent pas aux dimensions de l'interactivité ou du concept d'interaction développées dans la littérature que nous avons présentées précédemment (l'échelle ci-dessous présente par exemple des items comme la familiarité avec le Web ou la fréquence d'utilisation).

Tableau 1.3: Items de mesure de l'interaction sur le Web (adapté de Raman et Leckenby,1998)

1. How do you feel about the Web? Hate it very much :::::::: Love it very much
2. How frequently do you get on the Web ? Hardly ever :::::::: Get on frequently
3. How familiar are you with the Web? Not familiar at all :::::::: Very familiar
4. How much do you like the Web? Not at all :::::::: Very much
5. What would you say is your level of proficiency on the Web? Not proficient at all :::::::: Very proficient
6.What would you say your level of usage of the Web? Very little :::::::: Use It a lot

Pour étudier la nature et les conséquences de l'interactivité, Burgoon, Bonito, Bengtsson et Ramirez (1999/2000) décomposent l'interactivité en plusieurs attributs pertinents et testent empiriquement l'impact de ces attributs. Ils font tout d'abord l'hypothèse que l'interactivité est de valeur neutre, c'est à dire que la présence des caractéristiques de l'interactivité peut être bénéfique ou préjudiciable. Ils proposent une première conceptualisation de l'interactivité à partir de neuf éléments structurels :

1. la participation (le degré auquel les émetteurs et les récepteurs sont activement engagés dans l'interaction, contrairement aux monologues),
2. la médiation (si le format de la communication s'effectue au travers d'un medium ou non),
3. la contingence (le degré auquel les demandes et les réponses des individus dépendent des précédentes),
4. la richesse du media et de l'information (si le format utilise plusieurs modalités comme le texte, l'audio, la vidéo et le degré auquel la variété des supports présente l'information de manière riche ou pauvre),
5. la proximité géographique (si les utilisateurs sont physiquement au même endroit ou distribués),

6. la synchronisation (si l'interaction est en temps réel, s'il y a un feed-back bidirectionnel immédiat),
7. l'identification (le degré auquel les participants sont identifiés),
8. le parallélisme (si le format permet plusieurs communications de manière simultanée, des adresses multiples),
9. l'anthropomorphisme (le degré auquel les interfaces incorporent des caractéristiques humaines).

Selon ces auteurs, le deuxième moyen de conceptualiser l'interactivité est d'utiliser des expériences qualitatives vécues par les utilisateurs. Burgoon, Bonito, Bengtsson et Ramirez (1999/2000) mettent en avant trois éléments de l'interactivité à partir de l'expérience d'interactivité:

1. l'engagement dans l'interaction (le degré auquel les utilisateurs perçoivent qu'ils sont cognitivement, affectivement et comportementalement engagés dans l'interaction),
2. la mutualité (le degré auquel les utilisateurs perçoivent et créent des connexions relationnelles, de l'interdépendance, de la coordination),
3. l'individualisation (le degré auquel les individus perçoivent une proximité, l'identité des autres, une information personnalisée).

Ces auteurs testent ensuite l'impact de ces attributs grâce à plusieurs expériences. La spécificité de leur recherche repose sur le fait qu'ils analysent d'une part les interactions interpersonnelles et d'autre part les relations entre l'expérience de l'interactivité et ses résultats.

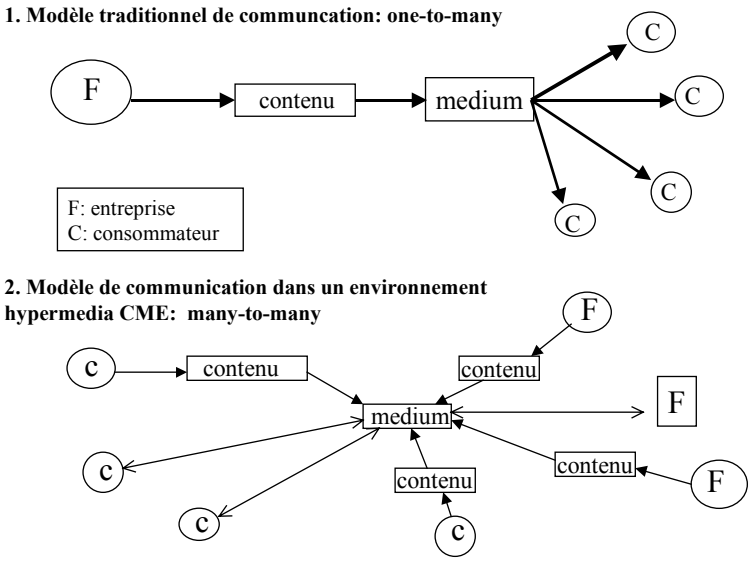
La thèse de Yuping Liu (2001) est consacrée à l'étude de l'interactivité et de ses implications sur le comportement du consommateur. L'auteur définit tout d'abord l'interactivité et ses mesures, développe une propre mesure de ce concept, puis analyse les implications de l'interactivité et enfin ses conséquences vis-à-vis d'un site et de la communication sur Internet. Dans un article publié un an plus tard, Liu et Shrum (2002) reprennent ces travaux et explorent les conditions selon lesquelles l'interactivité peut être bénéfique ou préjudiciable dans un contexte de communication. Soulignant tout d'abord le fait que personne ne sait réellement ce qu'est l'interactivité (dans les recherches académiques comme dans la presse managériale), ils affirment que l'interactivité est un construit complexe et multidimensionnel. Ils définissent l'interactivité comme le degré auquel deux groupes (ou plus) impliqués dans une communication peuvent agir l'un sur l'autre, sur le médium de communication, sur les messages et sur le niveau auquel ces influences sont synchronisées. Les deux auteurs proposent trois dimensions de l'interactivité:

1. le contrôle actif: besoin d'action constante, besoin de contribution, volonté qui influencent directement l'expérience de celui qui contrôle l'information,
2. la communication bidirectionnelle : capacité de communication réciproque entre entreprises- utilisateurs mais aussi entre utilisateurs-utilisateurs via Internet,
3. la synchronisation perçue: degré de simultanéité que l'utilisateur perçoit lors d'une communication entre le moment où il rentre ses informations et le moment où il reçoit la réponse.

La première dimension, le contrôle actif, est spécifique au domaine de l'Internet. Ce concept avait été introduit par Hoffman et Novak (1996) qui décrivent le Web comme un réseau de contenus reliés par des liens, constituant ainsi une structure parallèle et non linéaire. En contrôlant une telle structure non linéaire, les utilisateurs sont capables de personnaliser le flux d'information et de passer d'un lieu à un autre au sein du réseau, ceci étant totalement différent de la linéarité d'un médium comme la télévision. Comme le montre la figure 1.3 ci-dessous, Hoffman et Novak (1996) distinguent les modèles traditionnels de communication en marketing de ceux utilisés dans un environnement hypermédia CME (Computer-Mediated-Environment), défini par ces auteurs comme un réseau dynamique accessible par un ordinateur et qui permet aux consommateurs et aux entreprises de fournir et accéder à un contenu interactif (interactivité avec la machine) et de communiquer entre eux à travers ce médium (interactivité entre utilisateurs). Ils montrent donc que ces deux modèles diffèrent, comme le montre la figure 1.3. Le modèle traditionnel de communication (« one-to-many ») est utilisé classiquement en mass media : une entreprise (F) transmet un contenu à travers un médium à plusieurs consommateurs (C), tandis que le modèle de communication dans un environnement CME est un modèle « many-to-many » suggérant que la relation n'est pas entre l'émetteur et le récepteur mais plutôt entre l'environnement intermédiaire (l'ordinateur) et celui qui interagit avec cet environnement. Les consommateurs interagissent alors avec le médium (ils surfent sur le Web), tout comme les entreprises (en business-to-business par exemple). De plus, les entreprises et les consommateurs fournissent du contenu au médium, et les consommateurs peuvent interagir entre eux, tout comme les firmes peuvent interagir entre elles.

Figure 1.3: Modèles de communication dans un environnement traditionnel et un environnement hypermedia CME (Computer Mediated Environment)

(adapté de Hoffman and Novak, 1996)



La deuxième dimension de l'interactivité déterminée par Liu et Shrum (2002) est la communication bi-directionnelle. Les auteurs pour introduire ce concept s'appuient sur les théories d'Hoffman et Novak (1996) selon lesquelles Internet diffère d'un médium traditionnel par le fait que le Web permet aux consommateurs de donner leur feed-back, de communiquer entre eux, d'effectuer des transactions.

La troisième dimension introduite par Liu et Shrum pour définir l'interactivité est la synchronisation apportée par un médium comme Internet. Dans les media traditionnels, le temps de réponse à une demande est relativement long, alors que les communications peuvent être immédiates et synchronisées sur le Web.

Liu et Shrum (2002) proposent ensuite un modèle théorique pour tester les effets de ces dimensions individuelles de l'interactivité. Ce modèle s'appuie sur des théories en psychologie sociale, en psychologie cognitive et sur des recherches sur la personnalité. Les auteurs suggèrent plusieurs hypothèses selon lesquelles l'influence de l'interactivité sur l'efficacité d'une communication est fonction à la fois de la personne et de la situation.

Dans un article ultérieur, Liu (2003) développe une échelle pour mesurer l'interactivité de sites Web. D'autres échelles ont été proposées auparavant dans la littérature par Shankar, Smith et Rangaswamy (2000), Wu (1999), Cho et leckenby (1999) mais ces recherches selon Liu (2003) ne suivent pas le processus formel de développement d'une échelle et ne sont pas

claires sur plusieurs éléments (elles ne précisent pas par exemple le caractère unidimensionnel ou multidimensionnel de l'interactivité). Ces échelles sont également contaminées par des variables de réponse des utilisateurs (items de réponse affective ou d'intentions de comportement). Pour répondre à ces faiblesses, Liu (2003) développe l'échelle suivante (tableau 1.4) qui vérifie, selon son auteur, tous les critères de validité d'une échelle de mesure.

Tableau 1.4: Echelle de mesure de l'interactivité (Liu, 2003)

Items
Active control
I felt that I had a lot of control over my visiting experiences at this website.
While I was on the website, I could choose freely what I wanted to see.
While surfing the website, I had absolutely no control over what I can do on the site.
While surfing the website, my actions decided the kind of experiences I got.
Two-way communication
The website is effective in gathering visitors'feedback..
This website facilitates two-way-communication between the visitors and the site.
It is difficult to offer feedback to the website.
The website makes me feel it wants to listen to its visitors.
The website does not encourage visitors to talk back..
The website gives visitors the opportunity to talk back..
Synchronicity
The website processed my input very quickly.
Getting information from the website is very fast.
I was able to obtain the information I want without any delay.
When I clicked on the links, I felt I was getting instantaneous information.
The website was very slow in responding to my requests.

D'autres auteurs, McMillan et Hwang (2002) s'intéressent au concept d'interactivité perçue et développent une échelle de la mesure de ce concept, fondée sur trois éléments constitutifs de l'interactivité issus de leur analyse de la littérature : la bi-direction de la communication, le contrôle de l'utilisateur (sur la navigation, les choix), le temps (temps pour fournir ou trouver une information). Les items de l'échelle qu'ils proposent sont fournis dans le tableau 1.5 ci-dessous :

Tableau 1.5: Echelle de mesure de l'interactivité perçue (McMillan et Hwang, 2002)

Items
Enables two-way communication.
Enables concurrent communication.
Non concurrent communication.
Is interactive.
Primarily one-way communication.
Is interpersonal.
Enables conversation.
Loads fast.
Loads slow.
Operates at high speed.
Variety of content.
Keeps my attention.
Easy to find my way through the site.
Unmanageable.
Doesn't keep my attention.
Passive.
Immediate answers to questions.
Lacks content.

Dans une communication antérieure, McMillan (2000) s'intéresse aux relations entre les caractéristiques d'un site Web et l'interactivité perçue. Sur la base d'une expérimentation menée sur quatre sites, elle démontre que la présence de plus de caractéristiques interactives sur un site Web ne prédit pas une plus forte interactivité perçue. Sa conclusion est que l'interactivité doit donc d'abord résider dans l'œil du spectateur, montrant ainsi que les perceptions individuelles constituent un facteur important de l'interactivité.

Jee et Lee (2002) analysent les antécédents et les conséquences de l'interactivité perçue. Ils montrent que le besoin de connaissance et le niveau de compétences sur Internet sont des antécédents qui influencent la perception de l'interactivité d'un site Web lors d'une décision d'achat. En revanche, l'implication vis-à-vis du produit, le niveau d'expertise sur le produit, les défis sur Internet et l'expérience d'achat sur le Web n'apparaissent pas comme facteurs influençant la perception de l'interactivité du site. Ils montrent également que l'interactivité

perçue est très corrélée à l'attitude envers le site. En revanche, l'interactivité perçue ne donne pas une bonne prédiction des intentions d'achat.
Après avoir présenté les recherches concernant les facteurs constitutifs de l'interactivité, nous allons maintenant détailler la littérature marketing présentant l'interactivité comme un processus.

1.3.2.4 L'interactivité comme processus :

Un autre courant de recherche en marketing portant sur l'interactivité est l'ensemble des approches expérientielles analysant l'interactivité comme un processus. Ces approches ont été explorées par Holbrook et al. (1984) pour analyser l'expérience de consommation des individus. Hoffman et Novak (1996) introduisent pour mieux comprendre le comportement des utilisateurs d'Internet le concept de flux. Selon eux, le flux dans un environnement CME regroupe plusieurs composants essentiels de l'interaction du consommateur avec l'entreprise. Le premier auteur introduisant ce concept de flux est Csikszentmihalyi (1977) qui définit le flux comme un processus d'expérience optimale. Csikszentmihalyi et Lefebvre (1989) montrent que ce flux doit être précédé d'un ensemble de conditions nécessaires pour que l'expérience soit atteinte et qu'elle soit suivie d'un ensemble de conséquences correspondant aux résultats du processus. Appliquant ce concept de flux à Internet, Hoffman et Novak (1996) définissent l'expérience de flux dans un environnement CME comme un état intervenant durant la navigation sur le réseau, caractérisé par une séquence de réponses facilitée par l'interactivité de la machine, par le côté agréable de l'expérience , accompagné d'une perte de contrôle de soi et d'un renforcement de soi. Dans une expérience de flux, les consommateurs sont impliqués dans un acte de navigation dans lequel « plus rien ne compte » (Csikszentmihalyi (1990)). Hoffman et Novak (1996) soulignent le fait que les consommateurs focalisent leur attention sur l'interaction. Les conséquences d'une expérience sont alors l'apprentissage, l'exploration, les comportements participatifs, les expériences subjectives positives et un sens perçu du contrôle sur les interactions dans l'environnement CME. Le flux est donc un construit indispensable pour décrire les interactions homme-ordinateur (Csikszentmihalyi 1990). Hoffman et Novak (1996) utilisent les facteurs constitutifs de l'interactivité de Steuer (1992), précédemment décrits dans le modèle qu'ils proposent, pour représenter le dynamisme du processus de navigation dans un environnement CME.

D'autres auteurs, Bezjian-Avery, Calder et Iacobucci (1998), se sont aussi intéressés au concept d'interactivité comme processus, en comparant notamment la communication dans les nouveaux media interactifs par rapport à la communication traditionnelle. Ces auteurs définissent le marketing interactif comme un processus immédiatement itératif par lequel les désirs et besoins du consommateur sont d'abord non couverts, puis modifiés et satisfaits par le producteur. L'interactivité est alors un processus itératif entre l'entreprise et le consommateur, exigeant de l'information des deux parties, et tentant de faire correspondre intérêts et possibilités. Les itérations peuvent prendre un certain temps, permettant à l'entreprise de construire des bases de données pour proposer au consommateur une offre spécifique répondant à ses demandes. L'interactivité est fondamentalement dans ce cas la capacité à contrôler l'information. Tandis que, dans les media traditionnels, la présentation est linéaire et le consommateur passivement exposé à l'information produit, dans les communications interactives, les types d'informations vus par le consommateur dépendent de la manière dont celui-ci veut passer d'une étape à l'autre. Les auteurs distinguent plusieurs modes de « traversée » de l'information dans un système interactif : organisation en arbre hiérarchique avec des nœuds de décisions, organisation circulaire ou en réseau. Par une expérimentation utilisant le premier type d'organisation, les auteurs montrent que l'interactivité s'effectue au détriment de l'efficacité publicitaire, les individus dans un mode interactif portant peu d'attention au message, passant moins de temps à voir la publicité et ayant moins l'intention d'acheter les produits. Selon eux, l'efficacité du media interactif se mesure par l'engagement (temps passé par le consommateur) et la persuasion (préférences, intentions d'achat, affect positif). Enfin, les auteurs montrent que la communication interactive est plus efficace si le consommateur préfère l'information présentée visuellement plutôt que verbalement et si le contenu de la publicité est lui aussi plus visuel que verbal.

Après avoir présenté une revue de littérature sur le concept d'interactivité en marketing, nous allons aborder un autre courant de recherche fournissant une vision complémentaire sur les systèmes interactifs sur Internet : les approches en psychologie sociale.

1.4 L'INTERACTIVITE ENTRE INTERNAUTES ANALYSEE PAR DES APPROCHES EN PSYCHOLOGIE SOCIALE:

La deuxième source théorique que nous avons étudiée regroupe différents articles en marketing portant sur l'analyse de systèmes communautaires sur Internet dans lesquels les

internautes interagissent entre eux. Ces recherches portent majoritairement sur l'étude des communautés virtuelles sur Internet. Elles cherchent à comprendre le développement et le fonctionnement de ces communautés ainsi que les interactions entre leurs agents.

La notion de groupe communautaire est dans la littérature analysée par deux grandes approches théoriques : la première analyse le groupe d'un point de vue individuel dans des travaux issus de psychologie sociale ; la seconde étudie le groupe d'un point de vue holiste issu de théories en sociologie.

Dans cette partie, nous allons donc tout d'abord présenter les principales théories de psychologie sociale analysant les groupes sociaux, puis la littérature sur le marketing tribal étudiant les phénomènes de groupes au travers de théories en sociologie, avant de détailler les recherches sur les communautés virtuelles sur Internet, qui se fondent sur ces deux grands courants théoriques.

1.4.1 Les groupes sociaux en psychologie sociale :

Selon Serge Moscovici (1984), « la psychologie sociale est la science des phénomènes de l'idéologie (cognitions et représentations sociales) et des phénomènes de communication, et ce aux divers niveaux des rapports humains : rapports entre individus, entre individus et groupes, et entre groupes. ». Pour les psychosociologues, l'homme interagit avec les autres à travers des cadres sociaux et à travers un jeu de dépendances et d'influences. L'un des domaines historiques de la psychologie sociale est l'étude des groupes sociaux qui consiste à « appréhender un aspect des phénomènes sociaux à travers le mode d'insertion des individus et les conditions dans lesquelles s'expriment les interactions dans une telle structure » (Fischer, 1997, p.216). La définition d'un groupe social n'est pas univoque. Selon Darpy et Volle (1999), «tout rassemblement de personnes en un même lieu ne constitue pas un groupe ; ainsi, une file d'attente au guichet d'une banque ou les passagers d'un bus ne forment pas a priori un groupe. Un groupe est un ensemble d'individus qui sont liés entre eux, soit par des liens de proximité (comme la parenté ou l'amitié), soit par des liens symboliques (comme le partage de valeurs communes), soit les deux ». Ces groupes sociaux sont des groupes de référence qui ont une influence sur l'individu et sur son comportement d'achat. Une première distinction est menée en psychologie sociale entre les groupes primaires et les groupes secondaires (Fisher, 1997 ; Ladwein, 1999). Les groupes primaires sont des unités sociales dans lesquelles les individus ont des relations directes et personnelles avec un fort sentiment de cohésion. Ce sont par exemple la famille, les amis, les voisins. Les groupes secondaires

sont des organisations sociales dans lesquelles les relations sont moins directes et davantage déterminées par des codes, comme les associations ou les clubs. Fisher(1997) et Ladwein (1999) distinguent également en psychologie sociale les groupes informels des groupes formels. Dans les premiers, les rôles joués ne sont pas imposés et le type d'interaction ne repose pas sur une structure hiérarchique. En revanche, dans les groupes formels, les membres ont un rôle précis défini par une structure hiérarchique. Enfin, l'individu est influencé par son groupe d'appartenance (groupe auquel il appartient ou auquel on reconnaît qu'il appartient), dans lequel les interactions sont fonctionnelles ou émotionnelles, ou par un groupe de référence auquel il n'appartient pas (comme par exemple les champions sportifs, les vedettes de cinéma), mais dont il utilise les normes comme références de comportement. La nature de ces différents groupes sociaux est détaillée dans le tableau 1.6.

Tableau 1.6 : La nature des différents groupes sociaux (adapté de Fisher, 1997 et Ladwein, 1999)

Nature du groupe	Définition	Exemples
Groupe primaire	Unité sociale restreinte dans laquelle les individus ont des relations directes et personnelles, adhèrent aux valeurs qui leur sont proposées, et expriment un fort sentiment de cohésion	Famille, cercle d'amis
Groupe secondaire	Organisation sociale où les relations sont davantage déterminées par des codes et où les membres ont entre eux des relations plus ou moins imposées pendant la durée où ils sont ensemble (pas de relation directe ni personnelle)	Relations professionnelles, groupes associatifs
Groupe formel	Organisation définie : les membres ont une place assignée et des rôles prescrits, notamment par une structure hiérarchique	Organisation de défense professionnelle, syndicats
Groupe informel	Caractérisé par son émergence imprévue. Les membres qui le composent y sont de leur plein gré ; les rôles joués par chacun ne sont pas	Classe sociale, groupement idéologique

	imposés, et le type d'interaction ne repose pas sur une structure hiérarchique	
Groupe d'appartenance	Groupe auquel l'individu est reconnu appartenir (un groupe que l'individu pratique) ;caractéristiques principales : système d'interactions fonctionnelles ou émotionnelles, conformité au sein du groupe	Clubs, classe sociale
Groupe de référence	Réel ou virtuel, acceptation des normes du groupe et utilisation de celles-ci comme références perceptuelles ou comportementales	Champions sportifs, vedettes de cinéma.

La distinction groupe d'appartenance versus groupe de référence est celle qui a reçu le plus d'échos auprès des professionnels du marketing. En effet, les groupes auxquels un individu appartient (groupe d'appartenance) sont censés avoir une influence importante sur nos attitudes et nos valeurs, mais ne remplissent que partiellement leur rôle (Hyman, 1942, dans Fischer, 1997). Certains individus préfèrent adopter des normes de comportement de groupes auxquels ils n'appartiennent pas, mais de groupes auxquels ils aspirent. Ces groupes auxquels un individu se réfère (groupe de référence) jouent un double rôle de :

- fonction normative : il permet à l'individu d'évaluer son comportement au regard des normes du groupe,
- fonction comparative : il permet à l'individu de se positionner par rapport au groupe.

Plusieurs éléments durables structurant les groupes ont été identifiés en psychologie sociale :

- *la taille,*
- *le statut et le rôle des membres du groupe*: selon Fischer (1997, p.222), « le statut désigne la position objective occupée en fonction du niveau social ; il englobe un ensemble de caractéristiques objectives qui déterminent la place d'un individu sur une échelle sociale. Le rôle, quant à lui, peut être considéré comme l'aspect dynamique et subjectif du statut ; il désigne un modèle de conduite prescrite à un individu, lié aux exigences du statut et fonction des attentes du groupe ». Il existe un conflit de rôle au sein du groupe lorsqu'il y a divergence entre les attentes du groupe et le comportement réel de l'individu.

- *les normes du groupe* : « ensemble de règles élaborées par le groupe, mais qui se réfèrent de façon plus large aux normes sociales du milieu dans lequel il se trouve ; elles désignent le type de comportement jugé acceptable pour ses membres de sorte que chacun a tendance à s'y soumettre » (Fischer, 1997, p. 223).
- *l'activité du groupe* : les chercheurs en psychologie sociale ont montré qu'un groupe s'adonne à au moins deux activités : une activité de production (« accomplir une tâche ») et une activité de régulation (« créer ou maintenir les conditions psychologiques nécessaires au bon déroulement d'une tâche ou à la bonne entente entre ses membres ») (Fisher,1997, p. 224).

Selon Kotler & Dubois (2000), les groupes de référence influencent l'individu de trois façons : « ils proposent à l'individu des modèles de comportement et de mode de vie ; ensuite, ils influencent l'image qu'il se fait de lui-même ; enfin, ils engendrent des pressions en faveur d'une certaine conformité de comportement ». Fischer (1997) dresse un panorama des mécanismes d'influence des groupes sur l'individu. Selon lui, le groupe influence l'individu par tout d'abord ses *pressions normatives*. Les normes sociales contribuent à la normalisation des comportements qui aboutit à une conformité sociale. L'individu se rallie à l'opinion du groupe pour prendre sa décision. Ce phénomène a été illustré par des expériences célèbres (Fischer, 1997; Doise et al., 1991). Ce phénomène de conformité sociale peut prendre plusieurs formes : l'intériorisation du système proposé (l'individu abandonne ou change ses croyances propres pour adopter celles du groupe); l'identification (l'individu cherche à ressembler à ceux dont les attitudes lui paraissent désirables et adopte ainsi leurs attitudes); le suivisme (l'individu est conforme aux normes du groupe en public mais ils les rejettent en privé). Cette conformité est donc le résultat de pressions sociales, d'une recherche d'approbation sociale, ou encore de la construction sociale de l'objectivité (« notre jugement sur autrui et sur nous-mêmes se construit à partir des croyances qui, à leur tour, sont déterminées par les normes dominantes », Fischer, 1997, p.137). La deuxième source d'influence du groupe sur l'individu selon Fischer (1997) est la *persuasion* et la *manipulation.* Parallèlement aux normes du groupe, il existe un mécanisme par lequel un individu va chercher à obtenir d'une personne un changement d'attitude ou de comportement et ce de manière librement consentie. La *persuasion*, « moyen d'influence qui vise le changement d'attitude censé intervenir précisément comme un effet de la stratégie argumentaire » (Fischer, 1997, p.138), constitue l'une des formes de ce mécanisme. Les facteurs liés à la

personne, au message et au canal de communication modèrent l'efficacité du processus persuasif. La *manipulation,* selon Fischer (1997), est une action programmée, consciente, par laquelle celui qui l'utilise applique des connaissances psychologiques pour influencer d'autres personnes de telle sorte que ces dernières ne peuvent se rendre compte de la pression qui s'exerce sur elles. La troisième source d'influence du groupe sur l'individu est la *soumission à l'autorité*. Il s'agit de l'obéissance des membres du groupe à un pouvoir reconnu comme légitime. L'obéissance a fait l'objet de nombreux travaux, notamment ceux de Milgram (1965, 1974) soulignant le pouvoir d'influence d'une autorité, et la propension des individus à s'y soumettre (Fischer, 1997; Doise et al., 1991). Enfin, une quatrième source d'influence du groupe sur l'individu est distinguée par Fischer (1997): *l'action des minorités*. L'influence d'un groupe a été longtemps étudiée comme un processus majoritaire. Cependant, il est apparu qu'une minorité peut également exercer une influence sur une majorité, même si elle n'a pas, au départ, la crédibilité et l'autorité nécessaire à cet effet. L'influence minoritaire naît d'un conflit opposant cette minorité au reste du groupe. L'une des clés de l'influence minoritaire repose sur sa compétence : « c'est le caractère avéré d'une compétence qui va amener les membres d'un groupe à se soumettre à la position adoptée par la minorité. Ainsi, dans le cadre d'un groupe a priori non-hiérarchisé, l'influence minoritaire peut s'exercer de manière privilégiée, lorsque la minorité adopte une attitude consistante et lorsque la compétence de la minorité est avérée. L'influence minoritaire est durable en ce sens qu'elle affecte la norme du groupe » (Ladwein, 1999).

L'approche en psychologie sociale met également l'accent sur les phénomènes d'influence des membres en interaction dans un groupe, notamment les leaders d'opinion. Les recherches de ce courant sont liées à l'analyse de la structure hiérarchique des groupes qui a fait apparaître la capacité de certains individus à influencer le groupe : les leaders d'opinion. Fischer (1997, p. 240-241) distingue leadership, pouvoir et autorité. Selon lui, le pouvoir désigne la capacité de l'individu à exercer une contrainte sur autrui; cette capacité est rendue légitime quand la personne en question détient une autorité dans une structure hiérarchique. En revanche, le leader est celui qui exerce une influence prépondérante à l'intérieur d'un groupe donné, cette influence n'étant pas forcément liée à une position hiérarchique, mais plutôt au fait que ses positions soient acceptées par les autres qui auront par conséquent tendance à le suivre. Fischer (1997) distingue plusieurs types de leaders d'opinion, fondé soit sur les relations interpersonnelles, soit sur l'accomplissement de la tâche du groupe, selon qu'il est autoritaire, démocratique ou qu'il laisse faire. De nombreuses recherches en

marketing portent sur les leaders d'opinion, mais nous ne détaillerons pas ici ce courant afin de nous centrer sur le thème majeur de la thèse.

Plusieurs chercheurs en psychologie sociale se sont également intéressés à l'étude des phénomènes de bouche à oreille et de rumeurs. En effet, la dynamique de groupe est fondée sur une relation d'influence entre différents agents en interaction. De même, nous ne détaillerons pas ces recherches car elles sortent du champ d'investigation de la thèse.

En revanche, une source de littérature qu'il nous semble intéressant de détailler pour mieux comprendre les interactions entre différents individus est la théorie du marketing tribal, développé par Maffesoli et appliquée ensuite au marketing par Cova.

1.4.2 Le marketing tribal

Le marketing tribal étudie les phénomènes de groupes au travers de théories issues du domaine de la sociologie. En complément de l'approche psychosociale mettant l'accent sur les influences du groupe sur les interactions entre ses membres, l'approche tribale se place dans un contexte post-moderne, où la consommation sert de support au besoin de lien social des individus. Ce phénomène est lié au besoin croissant de lien social des individus à la fin du XXème siècle. Les institutions de socialisation traditionnelles ayant montré leur incapacité à l'intégration sociale des individus, on voit apparaître de nouvelles formes de socialisation au sein de structures éphémères et flexibles. Maffesoli (1996) parle de tribus, groupes d'individus qui partagent les mêmes centres d'intérêts ; des groupes peu structurés, au sein desquels les individus peuvent entrer et sortir librement, à la différence des communautés villageoises, la proximité et le contrôle social jouant un faible rôle dans ces groupes. Ceci correspond selon Maffesoli à l'émergence d'un mouvement inverse : la recherche du maintien ou de la création à nouveau de liens sociaux.

Développé à partir des théories de l'école Latine d'étude de la société, le marketing tribal, selon Cova, cherche à tempérer la vision psychosociale classique centrée sur les phénomènes d'influence sociale dans le groupe pour adopter une vision plus holiste qu'individualiste, centrée sur l'étude de ce qui fait le lien entre les individus (Cova & Roncaglio, 1999). Dans cette approche latine, la consommation est étudiée dans une perspective micro-sociale constituée d'interactions entre individus. Le leitmotiv est que le lien est plus important que le bien. Le courant post-moderne ouvert par Lyotard (1979), père fondateur, puis développé par des philosophes comme Lipovetsky et de sociologues comme Maffesoli postule que la modernité, avec ses grands récits dominants qui délimitent les systèmes de valeurs individuels

fondés sur la croyance en l'avenir, au projet et au progrès par le travail et centrés sur la raison a laissé la place, par un phénomène de saturation, à autre chose ; cette autre chose qui vient après la modernité, est donc post, elle lui succède et constitue une nouvelle direction après la précédente (Evrard et al., 2000). Dans cette société post-moderne, les individus sont guidés par des motivations hédoniques. Cova (1996) montre que cette vision post-moderne offre une alternative aux visions classiques du consommateur. Il propose « des leçons de marketing post-moderne ». L'objectif est de mobiliser les chercheurs et les professionnels du marketing autour de la resocialisation des individus. D'un côté, la montée de l'individualisme a conduit à la fragmentation de la société. Cette fragmentation, entretenue par le développement de l'industrie et du commerce, est l'une des plus visibles conséquences de l'individualisme qui règne dans nos sociétés. Face à cela, la post-modernité peut être comprise comme l'achèvement du processus de libération de l'individu, cette libération conduisant l'individu à la quête de son identité. Les produits et les services ont en effet, selon Cova, libéré progressivement l'individu des tâches aliénantes laissées par la tradition. En l'absence de référents sociaux, l'individu se tourne vers l'acte de consommation pour assouvir son besoin de lien social et pour se forger une identité, ce qui entraîne le développement des tribus reposant sur un lien consumériste. L'acte de consommation n'est pas un moyen de donner du sens à sa vie mais plutôt de nouer des liens avec les autres consommateurs : le lien est plus important que le bien.

Le mot tribu se réfère à la ré-émergence de valeurs quasi-archaïques comme le sens local d'identification, la religiosité, le syncrétisme, le narcissisme de groupe. Selon Cova (2002), le facteur commun aux tribus modernes est la communauté d'émotion ou de passion. Il reprend ici les idées de Maffesoli (1996) selon lequel les tribus postmodernes sont instables, de petite dimension, affectives et non déterminées par les paramètres établis de la société moderne. Le fait d'appartenir à ces tribus est devenu pour l'individu plus important que le fait d'appartenir à une classe sociale ou à un segment. Le statut social est progressivement remplacé par la position flexible et dynamique de l'individu au sein de ces tribus.

Les tribus diffèrent des groupes sociaux de référence présentés précédemment. Tout d'abord, selon Cova (2002), les individus peuvent appartenir à plusieurs néo-tribus. D'autre part, les tribus peuvent être simultanément des groupes primaires et secondaires. Comme dans les groupes primaires, les membres sont liés par des expériences concrètes de la vie quotidienne partagées. Mais ces tribus ne sont pas repliées sur elles-même car la condition de leur existence est d'interagir avec d'autres acteurs collectifs pour valoriser l'émotion partagée de leurs membres. Les tribus post-modernes diffèrent également des groupes de référence car

elles ne se focalisent pas sur les influences normatives du groupe ou sur l'influence des membres individuels du groupe entre eux. L'objectif premier est le lien qui garde et relie les individus.

Cova (2002) définit par conséquent la tribu moderne comme un réseau de personnes hétérogènes (en terme d'âge, sexe, revenu, etc…) qui sont liées par une passion ou une émotion partagée. Une tribu peut mener des actions collectives, ses membres ne sont pas de simples consommateurs, ils en sont aussi les avocats.

Cova (1996) fournit plusieurs explications à la croissance du tribalisme. Il souligne tout d'abord le *désir de communauté de nos contemporains*. Face à la dégradation des structures traditionnelles de lien social, les individus cherchent à nouer des relations, à recréer du lien affectif au travers d'un retour à la communauté. Ce désir de communauté s'inscrit dans des structures plus éphémères que les communautés traditionnelles, où l'individu est libre de partir quand il le souhaite. Le deuxième facteur explicatif de la montée des tribus est *l'imaginaire social de notre société*. Le désir de communauté ambiant s'est trouvé relayé par les media. Cova et Cova (2001) montrent que les consultants et autres publicitaires se sont approprié le mythe tribaliste pour les mettre en jeu dans la communication et le marketing. Ainsi, la société dans son ensemble est touchée par cet effet de mode, renforçant ainsi la tendance générale à la quête de lien social. Le troisième élément est la *consommation de plus en plus tribale*. On voit apparaître de nouvelles tribus de consommation partageant des valeurs émotionnelles autour d'objets totems. Les individus ne cherchent plus simplement à satisfaire des besoins utilitaires mais également à construire des expériences et du sens. Cette recherche de sens et d'expérience est relayée par la valeur de lien des produits et services au centre de la consommation. Enfin, Cova (1996, p. 94) souligne la *valeur de lien des produits et services.* Selon lui, « l'individu valorise les biens et services qui par leur valeur de lien permettent et facilitent l'interaction sociale, la co-présence physique, la rencontre. Les individus valorisent les rapprochements émotionnels autour d'objets consuméristes, d'objets totems».

Pour repérer une tribu, Cova et Roncaglio (1999) proposent un axe double. Tout d'abord, un axe *visible* qui se réfère aux pratiques et aux rites de la tribu. Ils opposent les pratiques occasionnelles (rassemblements ponctuels) aux pratiques institutionnelles (lieux institutionnels de rencontre). Ensuite, un axe *invisible* (latent), s'attachant à distinguer ce qui relève du vécu des membres (pratique quotidienne) et de l'imaginaire (où se trouvent les sympathisants de la tribu). Les communautés sont structurées autour de rites qui permettent le partage des valeurs communes et l'intégration de la tribu. Ces rites s'appuient sur des codes

propres à la communauté : les objets du culte (objets totems et associés) ; les costumes rituels (panoplie vestimentaire de la tribu) ; les lieux de culte; les images (idoles et icônes de la tribu). Les stratégies marketing issues du marketing tribal se fondent sur des bénéfices fonctionnels des produits qui doivent s'effacer aux profit de bénéfices « liants ». Les professionnels du marketing peuvent élaborer leur offre en veillant à offrir du lien plutôt que du bien. Cova et Cova (2001) proposent 10 étapes à mener dans une stratégie de marketing tribal, regroupée autour de cinq grand thèmes : l'étude ethnographique, le co-design, le mix tribal intensif, le mix tribal extensif et l'organisation marketing tribal, qui sont détaillés dans le tableau 2.7 ci-dessous.

Tableau 1.7 Les 10 étapes du marketing tribal (Cova & Cova, 2001)

Etude ethnographique: 1. Evaluer la valeur de lien de l'offre (rôle du produit/service dans la construction ou le maintien des liens interindividuels) 2. Repérer des regroupements tribaux autour d'une passion ou d'une pratique. 3. Mettre en évidence des rituels et autres codes implicites du mouvement tribal.
Co-design: 4. Prendre le temps de rendre "normale" la présence de la marque et de l'entreprise pour la tribu sans développer d'activité marchande. 5. Travailler en collaboration avec les membres de la tribu pour élaborer une offre capable de supporter les rituels tribaux.
Mix tribal intensif: 6. Offrir du lien plus que du bien. 7. Rester humble et garder une attitude non marchande.
Mix tribal extensif: 8. Mettre en jeu la tribu dans les campagnes de communication (valeur de légitimation du lien tribal de l'offre). 9. Fidéliser par référence affective à la tribu et à l'imaginaire tribal.
Organisation marketing tribal: 10. Gérer la dimension tribale du personnel de l'entreprise (recruter des membres de la tribu; reconnaître l'implication tribale).

Après avoir présenté les principales théories sur les groupes en psychologie sociale et les thèses en marketing concernant le marketing tribal, nous allons parcourir une revue de littérature plus spécifique portant sur l'application des théories précédentes pour mieux comprendre certains systèmes interactifs multi-agents sur Internet, comme les communautés virtuelles.

1.4.3 Les communautés virtuelles :

Le champ de recherche sur les théories de groupe appliquées à Internet s'est récemment étoffé par de nombreux articles portant sur les communautés virtuelles. Ce courant est fondé sur les approches en psychologie sociale et en sociologie que nous avons présentées dans les parties 1.3.1 et 1.3.2. En effet, Rheingold (1993) dans un ouvrage consacré aux communautés virtuelles affirme que l'utilisation de réseaux via les ordinateurs peut aider à restaurer des communautés qui n'existent plus dans la vie moderne. Il affirme que le cyberspace est l'un des moyens par lequel les individus peuvent trouver d'autres personnes possédant les mêmes centres d'intérêt au travers de discussions de groupes et d'interactions personnelles on-line et peut-être en les rencontrant dans la vie réelle. En marketing, les premiers auteurs qui se penchent sur ces communautés sont Armstrong et Hagel (1996), dont les travaux ont été largement cités par la suite. Pour ces auteurs, le Web ne se résume pas à un nouveau canal de distribution où l'entreprise peut nouer des relations personnalisées avec ses clients. Les entreprises ont la possibilité de créer des communautés virtuelles pour fidéliser leurs consommateurs. Ces communautés constituent également une source d'informations stratégiques pour l'entreprise. En connaissant ces communautés virtuelles, l'entreprise identifie quasi-instantanément les besoins de ses internautes. Armstrong et Hagel (1996) distinguent les types de besoins des consommateurs auxquels répondent les communautés virtuelles. Ils distinguent tout d'abord les *communautés de transaction*: elles facilitent l'achat et la vente de produits et de services qui fournissent de l'information liée à ces transactions. Les participants interagissent entre eux dans le cadre d'une transaction spécifique et reçoivent de l'information fournie par les autres membres de la communauté. Les auteurs illustrent cette catégorie de communauté par des exemples comme Virtual Vineyards, site Web de vente de vins de petits producteurs. Un autre type de communauté est constitué des *communautés d'intérêt* qui regroupent des participants interagissant sur des sujets spécifiques. Ces communautés supposent un degré plus élevé de communication interpersonnelle que les communications de transaction. GardenWeb, communauté de jardiniers qui échangent des informations sur ce thème, appartient à ce groupe. Les *communautés d'imagination* constituent la troisième catégorie. Les individus dans ces communautés créent de nouveaux environnements, des personnalités, des histoires, comme sur AmericaOnline où un participant créant son histoire s'invente une personnalité de baron médiéval. Enfin, les *communautés*

relationnelles regroupent des individus autour d'expériences de vie souvent intenses. Ceci peut aboutir à des connexions profondes et fortes entre personnes. Dans cette dernière catégorie de communautés, les individus connaissent souvent les identités des autres. A titre d'exemple, les auteurs citent Cancer Forum sur Compuserve, communauté apportant des informations et du support à des personnes atteintes de cancer et à leurs familles.

Armstrong et Hagel (1996) affirment que les communautés virtuelles auront des conséquences en marketing. Tout d'abord, les professionnels du marketing devront connaître les produits et services nécessaires à la communauté pour développer le réseau de ses membres. D'autre part, ils devront maîtriser et comprendre la technologie de manière à transformer en acheteurs des internautes qui récupèrent de l'information via des communautés. Les communautés virtuelles permettent également d'avoir une information détaillée sur les consommateurs que devront utiliser les équipes marketing. Enfin, les marketers doivent se demander qui organise les communautés et comment utiliser celles-ci pour renforcer les relations avec leurs clients. Enfin, les communautés affectent la nature même de certains produits, comme par exemple un magazine on-line : le produit est-il le magazine ou la communauté ?

Armstrong et Hagel (1996) énumèrent également un certain nombre d'éléments permettant de créer de la valeur dans les communautés. Tout d'abord, les communautés peuvent demander aux participants un droit d'entrée. Elles peuvent également faire payer une partie du contenu en cas de téléchargement. Des revenus peuvent également provenir de ventes et de publicités. Enfin, certaines communautés peuvent entraîner des synergies avec d'autres activités de l'entreprise.

Dans un ouvrage ultérieur, best-seller aux Etats-Unis, Hagel et Armstrong (1997) développent l'idée selon laquelle la réelle opportunité d'Internet est le développement des communautés virtuelles, qu'ils nomment Net Gain. Cette théorie a été reprise dans de nombreux articles en marketing. Cette théorie a suscité de nombreuses critiques de chercheurs et Hagel (1999) a publié un article ultérieur répondant aux réactions suscitées par son livre. L'idée de la théorie Net Gain est que les réseaux des communautés virtuelles apportent de nouvelles stratégies de business pour les entreprises. Que ce soient des communautés d'intérêts personnels (comme Motley Fool (www.fool.com), communauté autour des investissements en bourse), des communautés démographiques ou géographiques (comme www.parentsoup.com, rassemblant des parents ayant des enfants en bas âge), ou des communautés business-to-business (comme www.po.com, communauté de médecins), les communautés virtuelles rassemblent des individus qui ont un certain nombre de besoins ou des centres d'intérêts communs. La communication est donc un élément clef. La valeur est créée par l'intégration de contenu et

par des forums de discussion : plus les membres participent, plus la valeur qu'ils en retirent croît. Hagel (1999) affirme même que les communautés virtuelles constituent une nouvelle manière de faire croître son chiffre d'affaires. Au lieu d'avoir des vendeurs et de pousser des produits sur le marché, l'entreprise doit s'asseoir à côté du consommateur et l'aider à retirer la plus grande valeur des produits et services importants pour ce consommateur. L'auteur souligne que ce changement de culture est un réel défi pour les managers qui doivent comprendre ces processus de création de valeur. Il propose de développer des modèles de collaboration marketing dont l'objectif est d'accroître l'efficacité des consommateurs dans leurs achats de produits et services via ces communautés virtuelles. Ceci est rendu possible par le fait que ces communautés fournissent de nombreuses informations sur la manière dont les consommateurs utilisent les produits et les relient à d'autres aspects de leur vie. Hagel (1999) propose même d'utiliser les communautés virtuelles pour mener des focus groups 24h sur 24, 7 jours sur 7, afin de mieux connaître les attributs des produits susceptibles d'influencer les achats. Il affirme également qu'il faut capter et maintenir l'attention des consommateurs, ou encore développer des stratégies de sponsor sur ces communautés. Dans certains cas, le marché est même inversé : dans les communautés virtuelles, les vendeurs de l'entreprise peuvent être remplacés par les clients eux-mêmes qui au travers de la communauté vont trouver d'autres consommateurs.

Un autre type de recherche porte sur l'identification des membres des communautés virtuelles. Hagel et Armstrong (1999) distinguent quatre types d'internautes en fonction de leur contribution à la communauté.

- Les touristes : « classiquement, on entre dans une communauté virtuelle pour voir, en touriste » (Hagel et Armstrong (1999, p.78)). Certains resteront, d'autres partiront. Les touristes commencent à avoir une valeur économique pour la communauté lorsqu'ils passent à une autre catégorie. Ils peuvent jouer un rôle de rabatteur en faisant connaître favorablement la communauté dans d'autres forums ou sur un réseau, contribuant ainsi à attirer d'autres touristes, dont certains deviendront peut-être membres actifs.
- Les bâtisseurs : « il s'agit des membres les plus passionnés par la communauté et qui contribuent dans une mesure importante à lui apporter du contenu. Ils lui consacrent beaucoup de temps et, à bien des aspects, constituent le tissu conjonctif qui assure son unité » (Hagel et Armstrong (1999, p.78)). Leur valeur économique est double : ils passent beaucoup de temps dans la communauté, donc constituent des cibles intéressantes pour les annonceurs

(revenus publicitaires) ; ils contribuent également au contenu du forum, attirent et fidélisent les autres membres.

- Les utilisateurs : « parfois dits passifs, ce sont des gens qui passent plus de temps dans la communauté et qui bénéficient de ses informations sans beaucoup contribuer au contenu émanant des membres ni se comporter en acheteurs actifs » (Hagel et Armstrong (1999, p.79)). Ils peuvent constituer des cibles intéressantes pour les annonceurs (revenus publicitaires).
- Les acheteurs : il s'agit de membres qui achètent régulièrement des produits et des services via la communauté. Ces membres sont les plus précieux de la communauté.

Cette typologie relève moins d' une volonté de compréhension des membres que d'une volonté d'identifier les éléments les plus rentables pour l'entreprise. La finalité de Net Gain n'est pas principalement de comprendre les internautes mais de raccourcir le cycle de création d'une communauté de manière à la rendre plus profitable le plus tôt possible.

La performance des communautés virtuelles a fait l'objet d'autres recherches en marketing, comme celles de Bughin et Hagel (2000) qui s'intéressent à la création de valeur à long terme dans les communautés. Ils identifient les facteurs de performance des communautés on-line grâce à une approche par les coûts. La performance de la communauté est mesurée par trois critères : l'acquisition de visiteurs, la conversion du visiteur en consommateur et l'entretien du consommateur. Ils montrent que les sites de transaction sont moins capables de convertir du trafic en consommateur tandis que les communautés virtuelles combinent à la fois la croissance de trafic et la capacité de convertir les visites en nouveaux membres et en membres fidèles profitables. D'autre part, les communautés virtuelles accroissent leur efficacité avec l'âge. Elles ont donc une efficacité dynamique liée à leur cycle de vie. Enfin, les communautés virtuelles sont moins coûteuses que les sites de transaction. Les auteurs montrent ainsi l'intérêt des communautés virtuelles en tant que business model profitable sur Internet.

L'un des auteurs très cités dans les recherches en marketing sur les communautés virtuelles est Kozinets. Kozinets (1999) analyse les implications stratégiques liées au développement des communautés virtuelles et la participation dans la communauté. Se fondant sur les travaux de Komito (1998), Il définit les communautés virtuelles comme des groupes de personnes on line se comportant selon une même norme. Le terme communauté est utilisé dans son sens le plus fondamental comme un groupe d'individus qui interagissent socialement , qui développent des liens sociaux autour d'un espace commun. Dans les communautés sur

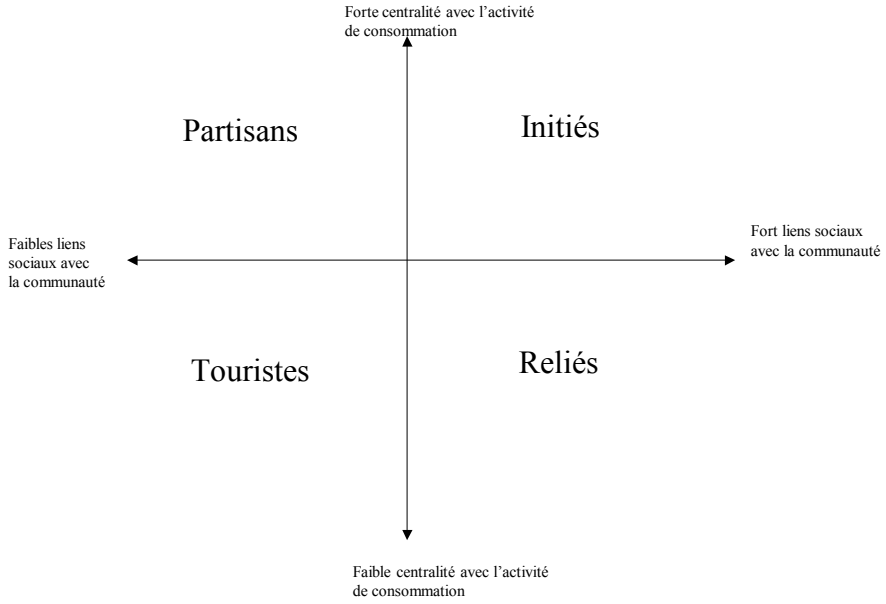
Internet, les interactions on-line constituent un « supplément » important au comportement social et au comportement de consommation. Les interactions on-line affectent également le comportement des individus en tant que citoyens, en tant que membre de la communauté et en tant que consommateurs.

Kozinets (1999) s'intéresse aux communautés de consommation, regroupant des individus intéressés par un thème lié à la consommation (comme par exemple la communauté autour des poupées Barbie). Selon l'auteur, la formation d'un membre d'une communauté de consommation dépend de deux facteurs indépendants. Tout d'abord, la relation que la personne a avec l'activité de consommation : plus l'activité de consommation est centrale pour l'individu, plus cet individu cherchera à valoriser le fait d'être membre de la communauté. Le deuxième facteur est l'intensité des relations sociales que l'individu a avec les autres membres de la communauté, les deux facteurs étant souvent liés entre eux.

Kozinets (1999) décrit l'évolution d'un membre d'une communauté virtuelle. Au départ, l'internaute va naviguer sur plusieurs pages Web pour lire des informations sur le sujet d'intérêt lié à la consommation. Par exemple, un internaute souhaitant acheter une automobile va d'abord visiter les sites officiels des constructeurs automobiles. Lorsque l'internaute est plus expert, il va visiter des sites fournissant un autre type d'information et va se mettre en contact avec d'autres consommateurs. Lisant les expériences des autres, il va aussi les interroger et devenir un participant occasionnel ou fréquent des groupes de discussion. Kozinets (1999) effectue une typologie des différents membres des communautés virtuelles et distingue quatre grands types de membres. Le premier groupe est constitué des touristes *(« tourists »)*. Ils n'ont pas de liens sociaux forts avec le groupe mais ont un intérêt superficiel ou passager vis-à-vis de l'activité de consommation. Viennent ensuite les « reliés » (traduction de *« minglers* ») qui ont de forts liens sociaux mais qui sont superficiellement intéressés par l'activité de consommation. Le troisième groupe est constitué de partisans (« *devotees* ») qui eux maintiennent un fort intérêt pour l'activité de consommation mais ont de faibles attachements sociaux vis-à-vis du groupe. Enfin, Kozinets (1999) isole un quatrième groupe : les initiés *(« insiders* ») qui ont de forts liens sociaux et un fort intéressement personnel dans l'activité de consommation.

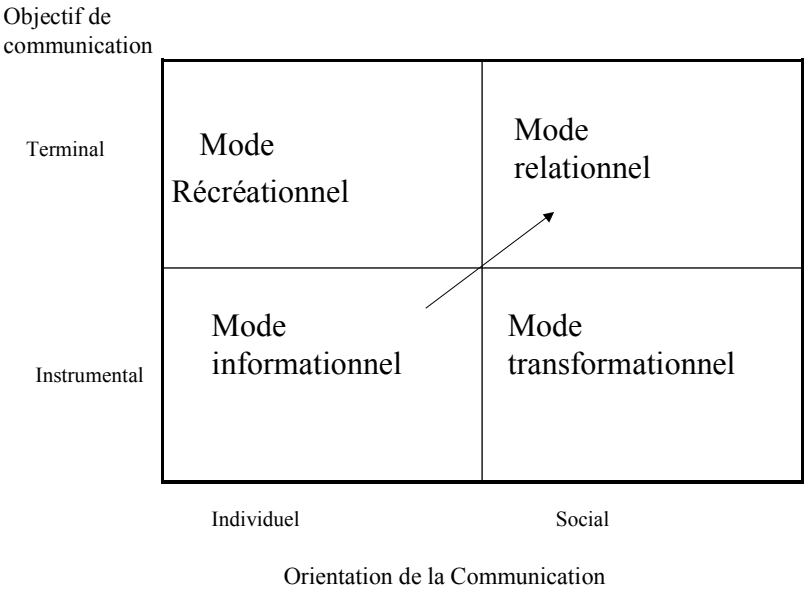
Cette typologie est représentée dans la figure 1.4 ci-dessous:

Figure 1.4: Les types de membres d'une communauté virtuelle de c onsommation (adapté de Kozinets, 1999)



Selon l'auteur, les partisans et les initiés constituent deux cibles stratégiques en marketing. Les touristes et les « reliés » quant à eux peuvent être socialisés et passer au stade supérieur de partisans ou d'initiés. Kozinets (1999) souligne également l'importance pour les équipes marketing du temps passé par le membre au sein de la communauté, ainsi que les modes d'interaction utilisés par le groupe. Si on comprend les modes d'interaction sociales utilisés dans les communautés de consommation, l'une des stratégies marketing possibles est d'adopter une segmentation basée sur ces interactions, comme par exemple les types d'interaction (informationnelle, relationnelle, récréationnelle, transformationnelle) qui indiquent précisément les potentiels pour obtenir les plus fortes réponses consommateur.
La figure 1.5 décrit ces différents modes d'interaction.

Figure 1.5: Les modes d'interaction dans les communautés virtuelles de consommation (adapté de Kozinets, 1999).



Les membres « reliés » ou initiés d'une communauté ont tendance à développer des interactions sociales et relationnelles dans leur communication. Pour eux, l'orientation sociale de la communication on-line va renforcer leur coopération et les standards du groupe (l'auteur cite à titre d'exemple les newsletter ou les membres qui écrivent les FAQ (Frequently Asked Questions)). L'objectif d'un mode d'interaction récréationnel est la satisfaction à court terme. Les « reliés » et les touristes utilisent majoritairement ce type de mode d'interaction car leurs liens sociaux restent superficiels. Dans le mode d'interaction de transformation, les consommateurs visent un lien social à long-terme. Ce mode est majoritairement utilisé par les « reliés » et les partisans dont l'intérêt pour l'activité de consommation les pousse à changer. Kozinets (1999) affirme qu'un marketing efficace vis-à-vis des communautés virtuelles est caractérisé par deux éléments : tout d'abord la tendance à créer des groupes homogènes, à utiliser une segmentation fondée sur la fragmentation des communautés en plusieurs sous-groupes (factions d'identité spécifique). Ensuite, l'activisme auquel les marketers devront s'adapter, fondé sur les interactions sociales on-line des membres de la communauté et sur leurs échanges d'information. Acquérant plus de pouvoir par cet échange d'informations, les consommateurs utilisent leurs activités on-line pour juger activement des offres proposées. Les entreprises devront donc être vigilantes de manière à maîtriser leur réputation et les messages envoyés au sein des communautés virtuelles. Kozinets (1999) souligne également le fait que les interactions basées sur de l'échange d'information transfèrent la connaissance et le

pouvoir des entreprises vers les consommateurs et que le marketing à l'âge de l'Internet devra apprendre à former des alliances avec ces communautés puissantes qui se développent on-line.

Dans une recherche ultérieure, Kozinets (2002) développe une technique d'analyse des communautés virtuelles qu'il nomme *Netnographie*. L'objectif est d'identifier et de comprendre les goûts, les désirs, les symboles, l'influence de certains consommateurs sur les prises de décision au sein de la communauté, les interactions entre ses membres. La Netnographie est une ethnographie des communautés virtuelles, nouvelle méthodologie qualitative adaptant les techniques d'ethnographie pour comprendre les cultures et les communautés dans les communications on-line. Cette technique est fondée sur l'information disponible dans les forums de discussions. Comparé à l'ethnographie traditionnelle, la Netnographie est plus élaborée, moins consommatrice de temps et moins intrusive. La limite de cette méthode, selon Kozinets (2002), repose sur la nécessité d'avoir un chercheur compétent qui interprète les résultats et sur le manque d'informations identifiant les membres des communautés on-line, rendant plus difficile une généralisation sur d'autres groupes. Kozinets (2002) décrit cinq types de communautés on-line susceptibles d'être analysées avec cette technique de Netnographie. Tout d'abord les forums (« *boards* »), nommés aussi « newsgroups », « usegroups » ou « usenet groups ». Ils sont organisés autour de produits ou de services particuliers, de styles de vie et sont liés à la consommation, donc particulièrement intéressants en marketing. Ce sont par exemple les communautés de McDonald's, de Sony ou concernant le ski. Le second type de communautés est constitué de pages Web reliées entre elles *(« rings* ») où les utilisateurs peuvent trouver des informations structurées et des détails d'achat sur un sujet spécifique. Un exemple est le site www.epinions.com dans lequel les consommateurs échangent des informations. La troisième catégorie concerne les listes de diffusion (« *lists* ») qui sont des listes d'adresses e-mail auxquelles sont adressées des informations sur un thème commun (une communauté professionnelle ou sur les jouets par exemple). Viennent ensuite les « *multiuser dungeons* », cercles constitués à l'origine autour du jeu donjons et dragons. Ces cercles se structurent autour de jeux en réseau et sont fortement récréationnels. La motivation des membres est également de rentrer en relation avec d'autres joueurs. Ces communautés regroupent souvent des consommateurs de produits technologiques ou de matériel informatique. La dernière catégorie correspond aux « *chat rooms* », espaces récréatifs mais avec de fortes interactions entre les membres. Ces communautés sont moins portées sur la consommation.

Kozinets (2002) propose donc via la Netnographie d'analyser les interactions entre les membres de ces communautés et le contenu des échanges en catégorisant les membres qui écrivent les messages selon les types de membres ci-dessus présentés (touristes, confondus, partisans, initiés). Il souligne également l'importance de l'interprétation du chercheur dans l'analyse des échanges au sein des communautés et la nécessité d'avoir une approche éthique lorsqu'on utilise la Netnographie qui doit rester une technique non intrusive.

En 2000, Brodin effectue une synthèse des recherches sur les communautés virtuelles en reprenant les approches de Kozinets et des autres auteurs pour explorer les relations que peuvent instaurer les communautés virtuelles. L'auteur montre qu'une meilleure connaissance de leur variété est un préalable à toute forme de marketing communautaire. Il propose trois stratégies aux entreprises : observer certaines communautés; établir un marketing relationnel avec des communautés virtuelles de consommation; établir un site Web communautaire.

D'autres auteurs ont étudié les interactions au sein des communautés virtuelles on-line. Williams et Cothrel (2000) s'intéressent à la manière de gérer les communautés. Ils analysent quatre organisations (Kaiser Permanente ; About.com ; Sun Microsystems ; Ford Motor co.) qui ont développé quinze communautés virtuelles et en tirent douze leçons fondamentales opérationnelles, décrites dans le tableau 1.8 ci-dessous.

Tableau 1.8: Les 12 leçons pour bien manager une communauté on-line (adapté de Williams & Cothrel, 2000)

1. Créer une masse critique de fonctionnalités
2. Collecter et utiliser le feedback des membres
3. Aménager le pouvoir de connexion personnelle
4. Amorcer la pompe avec de la communication
5. Soutenir les membres pour qu'ils s'entraident les uns les autres
6. Adapter les outils à la communauté
7. Jouer sur les motivations ds participants
8. Renforcer le cœur de la communauté
9. Fournir du matériel requis par la collaboration
10. Se concenter sur les communautés qui comptent
11. Former des communautés de personnes, non des applications
12. Développer le nombre de membres

Fischer, Bristor et Gainer (1996) définissent les communautés comme des groupes liés par des relations sociales et un sens de l'appartenance ou par un sujet d'intérêt commun et une identité partagée. Ces auteurs s'intéressent à des communautés on-line de consommation et mènent une étude empirique sur une communauté autour du thème de la dépression dans le

domaine de la santé. Ils montrent que les consommateurs acquièrent du pouvoir de en partageant de l'information entre membres et affirment que fournir un plus large accès à ces communautés de consommation serait bénéfique et que plus d'attention doit être portée sur ces communautés virtuelles pour accroître la valeur qu'elles procurent.

Moore, Mazvancheryl et Rego (1996) étudient la communauté virtuelle du jeu de Bolo et s'intéressent aux communautés liées aux jeux vidéo et aux produits high tech. Ils utilisent des techniques d'études qualitatives (entretiens en profondeur, observation des membres, contenu des pages Web) pour comprendre les interactions et les échanges entre les membres. Les auteurs analysent la consommation de loisirs sur la base de la théorie de la vue dramatique du monde (Celsi, Rose et Leigh (1993)) inhérente au développement des sociétés occidentales. La vue dramatique du monde repose sur l'interaction de forces sociales conflictuelles qui mènent à des tensions physiques et psychiques chez l'individu. Ces tensions menacent le contrôle de l'individu en induisant une perte d'autonomie. Ces forces conflictuelles construisent un stress physique et psychique qui doit se décharger pour résoudre le conflit. L'un des moyens de décharger ce stress est la consommation de loisirs. Moore, Mazvancheryl et Rego (1996) choisissent d'étudier le jeu de Bolo car il est multi-joueurs et très interactif. Après avoir décrit le profil des membres (plutôt des hommes entre 18 et 40 ans, étudiants ou professeurs), ils soulignent que l'un des principaux problèmes de la communauté est l'influx de nouveaux joueurs, découragés par les joueurs expérimentés qui s'opposent à eux. Ils mettent également en avant l'étiquette suivie par les joueurs et les sanctions attribuées à ceux qui ne suivent pas ces règles. La notion de flux et d'expérience où les membres retirent des bénéfices est particulièrement importante. Enfin, la recherche de soi est l'une des motivations pour devenir membre de la communauté et les compétences personnelles servent à définir le statut du membre.

Granitz et Ward (1996) étudient la nature de communautés virtuelles de groupes de discussion et sa relation à la communication sur le thème de la consommation. Analysant le contenu de plusieurs forums de discussions, ils montrent que 20% des mots sont dédiés à des recommandations produits, 20% à des discussions sur comment utiliser le produit, les autres catégories concernant la consommation du produit elle même et les décisions d'achat. Les auteurs montrent également l'influence de facteurs sociaux et cognitifs dans le contenu des communications : les experts du groupe jouent le rôle d'arbitre de l'opinion du groupe ; il y a une différence cognitive individuelle (expertise) liée à un rôle social (arbitre de l'opinion du groupe) distribuant la communication (direction des questions) et son contenu (critique ou non des autres). Les caractéristiques individuelles comme l'expertise, reliée à la structure

sociale du groupe de discussion, et les dimensions sociales influent sur la communication sur le thème de la consommation au sein du groupe. La structure sociocognitive influe aussi sur la communication et son contenu : les participants moins socialement proches du centre de la communauté posent plus de questions que le cœur de la communauté. Les acteurs centraux sont plus impliqués dans les réponses aux membres et exercent un pouvoir d'expert sur le groupe. Les marketers, selon Granitz et Ward (1996), doivent par conséquent se concentrer sur un petit nombre de participants qui influencent l'opinion de centaines d'autres.

Bagozzi et Dholakia (2002) conceptualisent la participation aux communautés virtuelles comme une action sociale intentionnelle et étudient le « nous-intentions » des membres. Selon les auteurs, les environnements digitaux prennent naissance dans les réseaux qui se développent sur la base d'une interaction sociale. Dans une communauté virtuelle, les objectifs mutuels peuvent être fonctionnels (par exemple l'échange d'informations concernant des produits), ou hédoniques (la création d'une expérience positive grâce à l'interaction) et la communauté joue le rôle d'un groupe de référence pour ses participants individuels. Se concentrant sur le rôle de l'individu et les influences sociales sur l'attraction de membres, Bagozzi et Dholakia (2002) adoptent une lentille de psychologie sociale en conceptualisant la participation des membres comme une action sociale intentionnelle. Les auteurs décrivent tout d'abord les attributs des communautés virtuelles : organisation autour d'un centre d'intérêt distinct qui constitue la raison d'être de la communauté ; conscience d'une « souche » (connexion intrinsèque avec les autres membres et sens collectif d'une séparation avec les non-membres) ; partage de règles d'interactions et d'un langage ; contenu créé par les membres de la communauté grâce à une participation active. Ils décrivent les différences des communautés virtuelles avec les communautés traditionnelles : les premières sont tout d'abord tirées par un choix volontaire (à la différence de communautés traditionnelles créées sur la base de critères géographiques ou de naissance par exemple) ; les membres peuvent en sortir facilement ; la participation est immersive et peut durer longtemps ; plus la participation est grande, plus les sources d'information sont importantes et influent sur les opinions et décisions d'achat. Bagozzi et Dholakia (2002) étudient ensuite les mécanismes psychologiques individuels et les processus d'influence sur la participation des membres. Se fondant sur les travaux de Tuomela (1995), ils définissent le « nous-intentions » comme l'engagement d'un individu dans une action de groupe, ce qui suppose l'accord explicite entre les participants dans cette action commune. Bagozzi et Dholakia (2002) utilisent le modèle de comportement vis-à-vis d'un objectif fixé (*Goal-Directed Behavior*) qui repose sur la théorie du comportement planifié selon laquelle l'intention personnelle de l'individu pour un type de

comportement est fonction de l'attitude de l'individu vis-à-vis de ce comportement (évaluation du comportement), de ses normes subjectives (pression sociale perçue pour développer ou non ce comportement) et du contrôle du comportement perçu (perception de la difficulté d'avoir ce comportement). Les auteurs montrent que les « nous-intentions » sont appropriées et doivent être mesurées par les marketers pour prédire la participation des membres et leur interaction. Ils montrent que la décision de participer à des communautés virtuelles, basée sur ces « nous-intentions », sont fonctions de trois antécédents : les émotions positives dans l'anticipation d'atteindre l'objectif de participation à la communauté ; l'identité sociale qui pousse les individus à participer ; les désirs qui ont une fonction de transformation et motivent la décision de participer à la communauté. Ils précisent que trois types de groupes sont pertinents à analyser en marketing dans le cadre de cette approche : les communautés on line où les individus se connaissent antérieurement à la création de la communauté (groupe d'amis par exemple) ; les groupes d'action coopérative et collective (comme les communautés autour de mailing lists) , où les membres coordonnent des actions individuelles ; les groupes d'action coopérative minimale, qui sont plutôt des groupes de co-action dans lesquels les individus ont un but commun mais pas d'intention mutuelle (comme par exemple les communautés autour du sport).

Dans un article récent se fondant sur leur recherche précédente, Dholakia, Bagozzi et Klein Pearo (2004) analysent les déterminants de la participation à une communauté virtuelle et analysent les antécédents et les variables médiatrices. Ils proposent également une typologie pertinente en marketing pour conceptualiser les communautés virtuelles et élaborer une distinction entre les réseaux et les communautés virtuelles basées sur de petits groupes d'individus. Pour les participants de communautés virtuelles basées sur des réseaux, la valeur du but de la communauté (définie comme la valeur dérivée de l'accomplissement de buts instrumentaux prédéterminés, comme par exemple recevoir ou envoyer de l'information) est un facteur clef de la participation. Les bénéfices sociaux (comme maintenir la connectivité interpersonnelle) et le renforcement social sont aussi des facteurs explicatifs de la participation. Les auteurs montrent également que le degré d'interactivité des processus de communication est influencé par la synchronicité de la communication (capacité qu'un participant formule et délivre une réponse en temps réel, autorisant ainsi un dialogue en temps réel) et par le nombre et la variété des entrées d'informations des participants. Le mode par lequel l'identité et l'information sur un membre sont communiqués aux autres membres est influencé par le type de communauté : dans les communautés à base de réseaux, la réputation

d'un membre est cruciale pour instaurer la confiance et développer des interactions sociales, contrairement aux communautés de petits groupes de membres.

Bickart et Schindler (2001) mènent une expérience dans laquelle les consommateurs doivent fournir des informations sur cinq sujets spécifiques liés à des produits en accédant à des discussions on-line. A la suite de 12 semaines d'expérience, les auteurs montrent que les consommateurs qui ont fourni de l'information ont un plus grand intérêt vis-à-vis du produit que les consommateurs qui ont reçu l'information de sources marketing traditionnelles. Leur expérience repose sur le fait que l'information dans les communautés est plus crédible que celle fournie par les marketers, les membres ayant plus confiance dans cette information plus persuasive. D'autre part, l'information échangée sur Internet est plus pertinente pour les consommateurs, car la perception d'une source similaire au récepteur influe sur l'influence persuasive accrue du message. Enfin, les forums sur Internet ont une grande capacité à générer de l'empathie parmi les lecteurs membres, pouvant ainsi influencer le comportement du consommateur vis-à-vis des bénéfices produit. Bickart et Schindler (2001) montrent ainsi que l'exposition à des forums en ligne génère plus d'intérêt pour la catégorie de produit que l'exposition à des sources d'information sur Internet générées par les marketers.

Plusieurs articles dans la littérature en marketing portent sur les communautés de marque, qui regroupent un univers plus large que les communautés virtuelles. Nous détaillerons ici l'article majeur de ce courant de recherche en raison de son lien avec les communautés virtuelles et de son importance au sein de la communauté des chercheurs en marketing. Muniz et O'Guinn (2001) définissent la communauté de marque comme une communauté spécialiste d'un produit ou d'un service, non limitée géographiquement, basée sur un ensemble structuré de relations sociales parmi les admirateurs d'une marque. Les auteurs définissent les caractéristiques de ces communautés : la conscience du type d'appartenance, ou encore la connexion intrinsèque que les membres ont les uns avec les autres et le sens collectif de la différence avec ceux ne participant pas à la communauté ; la présence de rites partagés et de traditions ; le sens de la responsabilité morale, sens perçu des obligations vis-à-vis de la communauté dans son ensemble et de ses participants. Utilisant ces critères dans leur étude empirique, Muniz et O'Guinn (2001) déterminent l'existence ou non de communautés de marque chez des habitants d'une petite ville nommée Fairlawn. Pour cela, ils analysent les interactions en face-à-face des participants mais aussi les interactions au sein de communautés virtuelles de marque créées sur Internet (pages Web dédiées à Saab, à Macintosh et à Bronco). Ils prouvent l'existence de communautés de marque dans les deux cas : en

face-à- face et dans un environnement avec ordinateur et en déduisent plusieurs implications sur la stratégie de marque des entreprises, notamment la nature sociale des marques.

Enfin, nous présentons un dernier article concernant les communautés virtuelles, issu d'un autre champ de recherche, celui des systèmes d'information. Weigand, De Moor et Van den Heuvel (2000) étudient les spécifications des systèmes nécessaires aux communauté virtuelles et montrent que les outils de communication modélisant des processus de workflow sont adaptés aux communautés virtuelles. Ils fondent leur recherche sur plusieurs courants théoriques et tout d'abord sur la théorie de rationalisation et de communication active d'Habermas (1981, 1990). Selon cette théorie, les individus ont de plus en plus des actions communicatives qui aboutissent à une coordination par des partages de connaissances et de normes non pas imposées mais volontairement acceptés par les participants dans le cadre d'une discussion ouverte. La communication active vise la vérité, l'exactitude des normes et la confiance et représente un champ théorique intéressant pour analyser les communautés virtuelles. D'autre part, plusieurs types de discours peuvent être utilisés dans les communautés virtuelles, mais le plus pertinent est un discours fondé sur des normes sociales valides, ou encore reconnues par la communauté. Ceci se traduit par un discours fondé sur un ensemble de règles : le discours doit être compréhensible pour tous les membres ; les participants doivent souhaiter atteindre un accord sur la base de la sincérité et de la responsabilité ; le processus de communication lui même repose sur la force du meilleur argument (aucun membre n'est exclu du discours). Le deuxième champ théorique utilisé dans cet article est la littérature sur les processus de communication et les échanges d'information (workflow). Les auteurs identifient plusieurs niveaux d'analyse : les transactions (unités composées d'actes de discours, de requêtes) regroupées dans des « boucles » de workflow (la « boucle » de workflow commence par une transaction ; après que l'exécutant ait exécuté l'action promise, la boucle est fermée par une conversation établissant que l'action a été bien accomplie); l'interaction, relation réciproque représentant deux boucles de workflow ; un ensemble d'interactions cohérentes liées entre elles, appelé scénario, représentant l'ensemble du processus. Ils proposent un modèle de système d'information (RENISYS) pour modéliser le processus des échanges au sein de la communauté virtuelle, intégrant tous les éléments ci-dessus présentés. Weigand, De Moor et Van den Heuvel (2000) montrent dans cette recherche que le design du système d'information d'une communauté virtuelle doit dépendre des caractéristiques uniques de la communauté que l'on peut analyser à partir des différents courants théoriques proposés par les auteurs et présentés ci-dessus.

Après avoir présenté ce courant spécifique de recherches en psychologie sociale qui éclaire par d'autres théories l'étude des systèmes interactifs multi-agents sur Internet, nous allons introduire maintenant une théorie développée en stratégie qu'il nous semble intéressant d'intégrer dans le cadre de cette thèse, car elle aborde le concept de création de valeur qui peut se développer dans un système grâce à l'interactivité entre plusieurs entités ou individus.

1.5 L'INTERACTIVITE EN STRATEGIE COMME SOURCE DE CREATION DE VALEUR

La troisième source théorique sur l'interactivité que nous allons présenter est un courant de recherche spécifique développé en stratégie sur l'interactivité et la co-production de valeur. Nous n'aborderons que les principaux éléments de cette théorie car celle-ci concerne un champ plus large que l'interactivité sur Internet qui reste notre sujet principal. En revanche, il nous semble intéressant d'introduire cette théorie susceptible d'ouvrir plusieurs pistes de recherche futures dans le domaine du marketing.

Dans un article précurseur de la Havard Business Review, Normann et Ramirez (1993) développent l'idée selon laquelle une stratégie interactive permet de passer de la chaîne de valeur à une constellation de valeurs. Selon les théories traditionnelles en stratégie, notamment celles de Porter, chaque entreprise occupe une position précise dans la chaîne de valeur, depuis les fournisseurs jusqu'au consommateur final en passant par l'entreprise elle-même. La stratégie consiste alors à bien positionner l'entreprise dans la chaîne de valeur, en définissant le bon marché, les bons produits, les segments de marché et les activités à valeur ajoutée que l'entreprise doit développer. Selon Normann et Ramirez (1993), un certain nombre d'éléments comme la concurrence globale, l'évolution des marchés et les nouvelles technologies modifient cette vision de la création de valeur : les entreprises ne doivent plus uniquement créer de la valeur ajoutée mais réinventer le système de création de valeur. L'idée sous-jacente de cette théorie est de reconfigurer les rôles et les relations entre ce que les auteurs nomment la constellation d'acteurs (fournisseurs, partenaires, consommateurs) de manière à ce que ceux-ci travaillent ensemble et co-produisent de la valeur. Pour illustrer cette théorie, Normann et Ramirez (1993) s'appuient sur l'exemple d'Ikea dont le succès repose sur une recomposition des rôles : sa stratégie de bas prix s'effectue en contrepartie d'un certain nombre de tâches prises en charge par le consommateur (livraison, montage des meubles). Dans les magasins, on fournit aux consommateurs des catalogues, des crayons, des bloc-notes de manière à ce qu'ils effectuent leur choix sans aide de vendeurs. Les auteurs affirment que

cette nouvelle logique de valeur entraîne trois conséquences sur la stratégie des entreprises. D'une part, dans un monde où la valeur n'est pas créée par une chaîne séquentielle mais par des constellations complexes, l'objectif d'une entreprise n'est pas de créer un produit de valeur pour le consommateur mais plutôt de mobiliser les consommateurs afin de créer de la valeur pour eux. D'autre part, dans des systèmes créateurs de valeur, les offres les plus attractives englobent les consommateurs, les fournisseurs et les partenaires dans de nouvelles combinaisons. La stratégie de l'entreprise consiste à reconfigurer ces relations. Enfin, si la clef pour créer de la valeur est de co-produire des offres qui mobilisent les consommateurs, la seule vraie source d'avantage compétitif est la capacité à concevoir un système entier de création de valeur qui fonctionne. Ceci signifie que l'entreprise doit définir le système, mobiliser et former les acteurs, laisser le consommateur arbitre final du succès ou de l'échec de cette stratégie. Normann et Ramirez (1993) citent deux exemples de réussite d'une telle stratégie: un réseau de 300 pharmacies danoises qui en plus de leur magasin ont développé un vrai centre d'information et de services autour de la santé; La Générale et la Lyonnaise des eaux en France qui ont reconfiguré le rôle des différents acteurs de leur système en créant, acquérant ou s'alliant avec des entreprises possédant des compétences spécifiques. Le secret de cette nouvelle création de valeur repose sur des relations et une connaissance de plus en plus poussées du système dans lequel se trouve l'entreprise. Cette stratégie interactive entre tous les acteurs du système permet donc, selon Normann et Ramirez (1993), de créer une constellation de valeur et non plus une chaîne de valeur, et de développer par conséquent des stratégies à succès dans le contexte actuel de mondialisation et de développement des nouvelles technologies.

Dans un article ultérieur, Ramirez (1999) approfondit cette théorie et développe l'idée selon laquelle les consommateurs peuvent créer de la valeur, peuvent co-produire et co-créer avec les producteurs, au lieu de détruire la valeur produite par ceux-ci comme dans les cadres conceptuels précédents. Après avoir présenté un panorama du concept de co-production de valeur sur les deux cent quatre-vingt-dix dernières années, il analyse les implications de ce concept sur le plan pratique et théorique dans le cadre de recherches futures. Ramirez (1999) s'attache notamment aux implications de la co-production de valeur sur la définition des domaines stratégiques des entreprises, sur l'organisation du travail, sur le management et sur la transition vers une économie co-productive (cf Tableau 1.9).

Tableau 1.9 : Co-production de valeur: origines intellectuelles et implications pour les managers et les chercheurs (Adapté de Rafael Ramirez (1999, Strategic Management Journal)).

Vision industrielle	**Vision co-production**
La création de valeur est séquentielle, unidirectionnelle, mieux décrite par les chaînes de valeur	La création de valeur est synchrone, interactive, mieux décrite dans les «constellations de valeur »
Toutes les valeurs peuvent être mesurées en termes monétaires	Certaines valeurs ne peuvent être mesurées ou monétisées
La valeur est additionnée	Les valeurs sont co-inventées, combinées et reconciliées
La valeur est une fonction d'utilité et de rareté	L'échange est source d'utilité et de rareté
Les valeurs sont objectives (échange) et subjectives (utilité)	Les valeurs sont contingentes et actuelles (établies de manière interactive)
Les consommateurs détruisent la valeur	Les consommateurs co-créent des valeurs
La valeur est réalisée à la transaction, seulement pour le fournisseur (événement)	La valeur est co-produite, avec le consommateur, dans le temps, pour les deux co-producteurs (relation)
Le modèle des 3 secteurs (primaire, secondaire, tertiaire) est pertinent	Le modèle des 3 secteurs n'est plus pertinent
Les services sont des activités séparées	Les services constituent un cadre pour toutes les activités considérées comme co-produites
La consommation n'est pas un facteur de production	Les consommateurs sont des facteurs de production (actif)
Les acteurs économiques ont analysé la propriété comme principal acteur à un instant donné	Les acteurs économiques ont analysé la propriété comme ayant plusieurs rôles différents simultanément
L'entreprise et l'activité sont les unités d'analyse	Les interactions (offerings) sont les unités d'analyse

Dans un ouvrage dédié à leur théorie sur la stratégie interactive, Norman et Ramirez (1994) décrivent un système de constellations de valeur fondé sur de multiple interactions entre les

acteurs, montrant que les entreprises ne gagnent pas de l'argent grâce à leurs consommateurs (ou leurs produits) mais plutôt grâce à des activités de création de valeur de leurs consommateurs qui ne sont possible que s'il y a interaction. Selon ces auteurs, les consommateurs s'engagent dans des activités dans le but d'obtenir de la valeur, non seulement financière mais aussi sociale, psychologique, esthétique, morale. Le consommateur n'est pas uniquement une contrepartie commerciale, il dépend de la valeur qu'il crée. Leur cadre théorique décrit les caractéristiques de co-production, de dialogue et les caractéristiques des formes émergentes de systèmes de création de valeur. Selon cette théorie, l'interactivité est par conséquent un mécanisme permettant aux entreprises et à leurs consommateurs de co-produire de la valeur.

Après la présentation de cette théorie spécifique sur l'interaction et la co-production de valeur développée en stratégie, nous allons maintenant présenter un courant de recherche issu d'autres domaines, l'informatique, l'intelligence articielle et la recherche opérationnelle, qui porte sur les systèmes multi-agents. Ce courant nous apporte de nombreux éclairages par rapport à la problématique de la thèse qui est d'étudier et de modéliser les systèmes multi-agents interactifs sur Internet.

1.6 LES PROCESSUS INTERACTIFS DANS LES SYSTEMES MULTI-AGENTS :

La quatrième source théorique à la base de cette thèse concerne le courant de recherches issu de l'intelligence artificielle et de l'informatique concernant les systèmes multi-agents. Les systèmes multi-agents (SMA) sont apparus aux Etats-Unis à la fin des années 1970, l'une des premières applications étant le développement d'un logiciel de reconnaissance vocale. Depuis, ce domaine a connu un essor considérable et aujourd'hui les recherches s'orientent vers des applications de plus en plus élaborées. Les systèmes multi-agents sont utilisés dans plusieurs domaines d'applications (Ferber, 1997), en particulier la robotique, la résolution distribuée de problèmes, la modélisation et la simulation des systèmes complexes. Plusieurs raisons expliquent leur développement. Tout d'abord, l'approche multi-agents est issue de recherches en Intelligence Artificielle (IA) qui cherchent à modéliser et développer des systèmes intelligents. Ces recherches en Intelligence Artificielle se sont cependant heurtées à de nombreux problèmes théoriques, notamment l'explosion combinatoire dans la manipulation des connaissances. Selon Mandiau, Grislin-Le Strugeon et Péninou (2002), les systèmes multi-agents ont permis de résoudre ce problème en évitant de travailler sur une seule entité et en répartissant la manipulation des connaissances sur plusieurs entités

intelligentes. D'autre part, les systèmes multi-agents ont aussi permis de décrire la coopération homme-ordinateur qui est le thème de nombreuses recherches cherchant à étudier les systèmes homme-machine, environnements où interagissent hommes et ordinateurs. Enfin, l'essor des SMA est également lié aux évolutions technologiques de l'informatique, au développement du commerce électronique (où de nombreuses interactions existent entre plusieurs entités intelligentes d'objectifs différents), à l'utilisation d'Internet et au nombre croissant de ses utilisateurs. Mandiau, Grislin-Le Strugeon et Péninou (2002) affirment que par conséquent la modélisation des interactions entre systèmes intelligents, entre individus et machines, devient un besoin important, et que l'étude des interactions étant le fondement même des SMA, ces modèles répondent donc à ce besoin. Internet en est une application évidente. Alors que les premiers systèmes d'ordinateurs étaient des entités isolées, communiquant seulement avec leur opérateur humain, les systèmes aujourd'hui sont interconnectés. Ils sont souvent reliés au réseau Internet. Selon Wooldridge (2002), les chercheurs et les professionnels en entreprise doivent par conséquent s'adapter à ces nouveaux environnements afin de mieux comprendre et appréhender les processus d'interactions entre tous les agents sur Internet.

Selon Ferber (1995), les recherches sur les systèmes multi-agents suscitent un certain nombre d'interrogations : qu'est-ce qu'un agent qui interagit avec d'autres agents ? Comment peuvent-ils coopérer ? Quels sont les modes de communication nécessaires pour qu'ils se répartissent des tâches et coordonnent leurs actions ? Quelle architecture peut-on leur donner pour qu'ils satisfassent leurs buts ? Le même auteur souligne que les SMA permettent aussi de modéliser des comportements naturels, d'étudier l'émergence de phénomènes complexes et de mieux comprendre ce qu'est une « intelligence collective ». Les SMA se distinguent des approches systémiques plus classiques en considérant que l'action et l'interaction sont les éléments moteurs de la structuration d'un système dans son ensemble (Ferber, 1995). Ferber (1995) définit une interaction comme une mise en relation dynamique de deux ou plusieurs agents par le biais d'actions réciproques. Les interactions s'expriment ainsi à partir d'une série d'actions dont les conséquences exercent en retour une influence sur le comportement futur des agents. Les agents interagissent le long d'une suite d'évènements pendant lesquels les agents sont d'une certaine manière en contact les uns avec les autres, que ce contact soit direct ou qu'il s'effectue par l'intermédiaire d'un autre agent ou de l'environnement. Les interactions sont non seulement la conséquence d'actions effectuées par plusieurs agents en même temps, mais aussi l'élément nécessaire à la constitution d'organisations sociales. C'est

pourquoi, selon Ferber (1995), il est impossible d'analyser des organisations sociales sans tenir compte des interactions entre leurs membres.

Plusieurs recherches portent sur la définition d'un système multi-agent, définition reconnue et reprise par la plupart des chercheurs dans ce domaine.

Les systèmes multi-agents sont des systèmes composés de plusieurs éléments qui interagissent entre eux, nommés agents (Wooldridge, 2002). Ferber (1997) définit un agent comme une entité qui :

- se trouve dans un système ouvert comprenant un ensemble d'applications, de réseaux de systèmes hétérogènes,
- peut communiquer avec d'autres agents,
- est mue par un ensemble d'objectifs propres (on parle alors d'agent intentionnel),
- possède ses propres ressources,
- ne dispose que d'une représentation partielle des autres agents,
- possède des compétences (services) qu'elle peut offrir aux autres agents,
- a un comportement tendant à satisfaire ses objectifs, en tenant compte d'une part des ressources et des compétences dont elle dispose, et d'autre part de ses propres représentations et des communications qu'elle reçoit.

Dans un SMA, les agents ont deux caractéristiques fondamentales. Tout d'abord, ils sont capables d'actions autonomes, c'est à dire qu'ils décident eux-même des actions nécessaires pour atteindre leurs objectifs. D'autre part, les agents sont capables d'interagir avec d'autres agents. Pour cela, ils nécessitent des capacités de coopération, de coordination, de négociation qu'ils appliquent dans leur vie quotidienne pour interagir efficacement avec les autres agents. Deux grands courants de recherches sur les SMA se distinguent (Wooldridge, 2002). Le premier concerne les agents individuels : les agents sont en très grand nombre et constituent un ensemble dont la structure est simple. Le deuxième s'intéresse à la structure organisationnelle des agents. Cette structure est plus ou moins flexible et le comportement des agents est délimité par leur rôle respectif. Par rapport à ces deux grandes tendances, les grandes questions auxquelles répondent les recherches sur les systèmes multi-agents sont les suivantes (Wooldridge, 2002) :

- Comment construire des agents indépendants et capables d'actions autonomes afin qu'ils prennent en charge les tâches qu'on leur attribue ?
- Comment construire des agents capables d'interactions (coopération, coordination et négociation) afin qu'ils prennent en charge les tâches qu'on

leur attribue, en particulier lorsque les autres agents n'ont pas les mêmes objectifs ?

Les chercheurs en SMA se sont surtout intéressé au départ aux systèmes informatiques et à la robotique. Aujourd'hui, les recherches sur les SMA portent aussi sur la capacité sociale des agents, et d'autres questions sont étudiées dans ces recherches (Wooldridge, 2002) :

- Comment développer la coopération dans des sociétés d'agents à objectifs individuels ?
- Quelles sont les langages que peuvent utiliser les agents pour communiquer aux autres agents leurs croyances et leurs aspirations ?
- Comment des agents ayant leur propres objectifs individuels se reconnaissent-ils lorsque leurs croyances, leurs objectifs et leurs actions rentrent en conflit ? Comment peuvent-ils parvenir à un accord ?
- Comment les agents autonomes coordonnent-ils leurs activités pour atteindre des objectifs en coopération avec les autres agents ?

On peut remarquer ici que les problématiques soulevées dans les systèmes multi-agents sont proches de questions traitées en sciences sociales, notre domaine de recherche en marketing.

La littérature sur les systèmes multi-agents est vaste et touche de nombreuses problématiques. Afin d'être cohérent avec le sujet de la thèse, nous avons sélectionné, pour cette revue de littérature sur les SMA, les articles selon deux critères : ces articles portent sur des systèmes multi-agents sur Internet ; ils touchent des problématiques intéressantes par rapport au domaine du marketing.

L'étude de certains phénomènes sur Internet, notamment les agents virtuels, par une approche SMA est développée par Wooldridge (2002). Selon lui, Le Web présente d'énormes potentiels en tant que ressource d'information mais pose également plusieurs problèmes, comme la surcharge d'informations. Les agents sur le Web apportent une solution à ce type de problème. Un agent d'information a accès à plusieurs sources d'informations et est capable de manipuler les informations de ces sources pour répondre à des questions posées par les utilisateurs. Le réseau de ces sources d'information constituent des systèmes d'information coopératifs et intelligents (Papazoglou et al., 1992). Les sources d'information peuvent être de différents types comme les bases de données traditionnelles. Trouver une solution à la demande d'un utilisateur suppose qu'un agent accède à l'information au travers d'un réseau. Un scénario typique est un agent qui recherche de l'information sur des voitures : il va consulter différents sites web et va récupérer ensuite plusieurs contacts de concessionnaires

ou de conseillers de clientèles. Le Web est particulièrement efficace dans ce cas pour plusieurs raisons. Il permet tout d'abord d'accéder à un réseau de ressources d'informations ; il fournit une interface uniforme de ressources multi-media incluant du texte, des images, du son, de la vidéo; il est basé sur de l'hypertexte permettant de relier des documents entre eux ; il est simple à utiliser. Cependant, selon Wooldridge (2002), le Web pose deux types de problèmes : ceux liés à des facteurs humains d'une part; ceux liés à des facteurs organisationnels d'autre part. Maes (1994) montre que la difficulté pour un utilisateur du Web est le problème de la surcharge d'information. Les individus sont submergés d'informations et il est difficile pour eux de filtrer la bonne information ou les informations les plus importantes. Les moteurs de recherche comme Google ou Yahoo, en indexant et structurant l'information, ont permis de diminuer ce problème, mais ils manquent de fonctionnalités : ils ne répondent pas aux demandes spécifiques d'un individu donné et ne sont basés que sur de l'hypertexte et non sur les supports video ou sons du Web. Par conséquent, la recherche d'informations sur le Web reste encore un problème. De plus, les individus sont souvent perdus ou s'ennuient lorsqu'ils naviguent sur des sites Web (Wooldridge, 2002). Lorsque l'internaute recherche de l'information, il est facile pour lui de la rater ou de mal comprendre des choses. La plupart des pages Web essaient d'être attractives et très animées pour attirer les internautes, mais il y a un paradoxe entre cette animation, ce divertissement et l'objectif de fournir une information précise, surtout lorsque celle-ci est complexe.

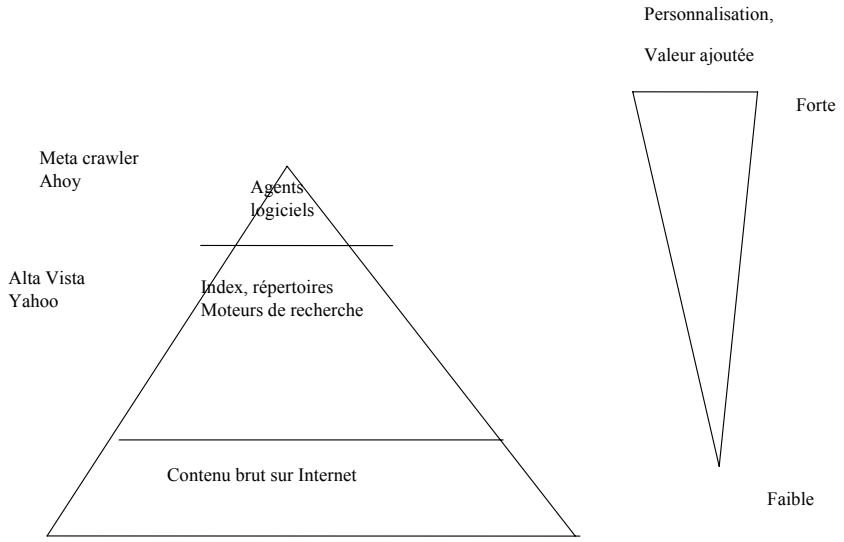
Le deuxième problème que pose Internet est lié à des facteurs organisationnels. A part le format HTML, il n'y a pas vraiment de standard sur la manière dont doit être constituée une page Web. De plus, fournir du contenu reste coûteux.

Par conséquent, Wooldridge (2002) affirme que des outils sont nécessaires pour répondre à ces problèmes, pour mieux comprendre Internet et développer son potentiel.

Montrant que l'approche multi-agents fournit ces réponses, de nombreux chercheurs ont développé des agents d'information personnels, comme Pattie Maes du MIT qui a créé en 1994 un assistant pour email (MAXIMS), permettant de prioriser, supprimer, classer et archiver des mails. Le principe de fonctionnement est le suivant : pour chaque événement arrivant sur le système (comme par exemple l'arrivée d'un e-mail), l'agent MAXIMS l'enregistre sous la forme situation/action. La situation est caractérisée par les attributs de l'événement (par exemple l'expéditeur de l'e-mail, les destinataires, le contenu du message, les mots-clés). Lorsqu'une nouvelle situation arrive, MAXIMS la compare aux situations précédentes et aux règles de l'utilisateur fixées lors des situations précédentes. Ensuite, à partir de ces règles, MAXIMS prédit ce que l'utilisateur va faire et agit en conséquence.

Etzioni et Weld (1995) identifient les différents types d'agents sur le Web qui selon eux vont apparaître dans le futur. La première catégorie est constituée de *guides de voyages*. L'objectif de ces agents est d'aider l'internaute à répondre à la question « où vais-je ensuite ? » lorsqu'il navigue sur Internet. Ces agents apprennent les préférences des utilisateurs, comme dans le cas de MAXIMS, et plutôt que de fournir un hyper lien simple, ils indiquent l'intérêt de ce lien. Le deuxième type d'agents est l'*agent d'indexation*. L'idée est d'utiliser l'information de base fournie par les moteurs de recherche comme Google, en ajoutant la connaissance des préférences et des objectifs des utilisateurs, de manière à fournir un service personnalisé, comme le montre la figure 1.5 ci-dessous.

Figure 1.5: La chaîne alimentaire d'information du Web (adapté d'Etzioni , 1996)

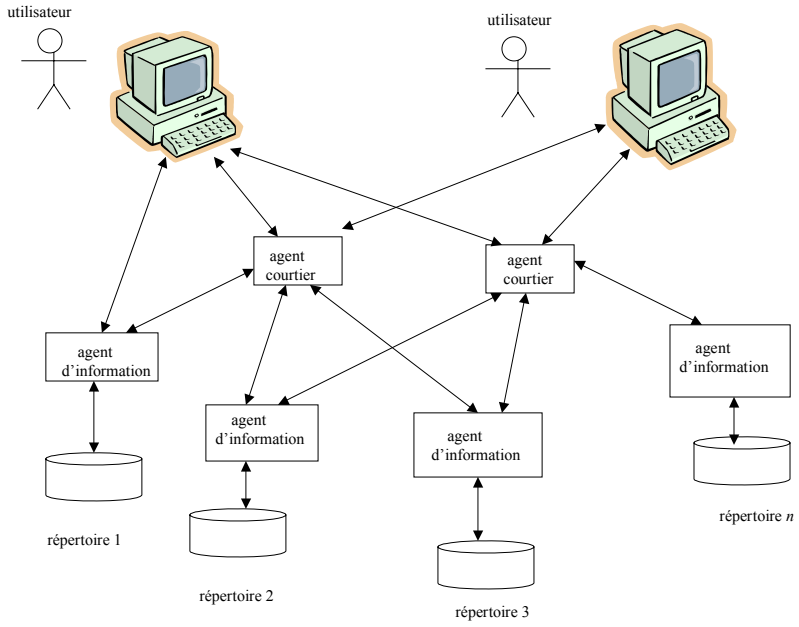


La troisième catégorie d'agents sur Internet identifiés par Etzioni et Weld (1995) sont les *détecteurs de FAQ* (Frequently Asked Questions). L'objectif de ces agents est de diriger les utilisateurs vers les FAQ pour que les internautes obtiennent une réponse à leurs questions. La dernière catégorie sont les *détecteurs d'expertise*. Lorsque l'internaute rentre des mots clefs dans un moteur de recherche (comme Google), le moteur prend simplement chacun des mots clefs et effectue une recherche sur eux. Mais les moteurs de recherche ne savent pas ce que l'internaute signifie par cette recherche ou ce qu'il veut réellement. Les détecteurs d'expertise cherchent à comprendre ce que l'internaute veut et le contenu des informations fournies pour fournir un meilleur service final.

Etzioni (1996) propose un modèle d'agents d'information qui apporte de la valeur ajoutée à l'information fournie sur le Web, qu'il nomme la chaîne alimentaire d'information (cf. figure 1.5) : le plus bas niveau d'information est l'information brute (les pages Web des entreprises ou les pages personnelles). Le niveau supérieur est l'ensemble des services qui consomment ce contenu brut, comme les moteurs de recherche de type Google, Yahoo ou Lycos. Etzioni propose un troisième niveau, en développant des méta-moteurs de recherche permettant de personnaliser les services.

Un autre courant de recherche sur les systèmes multi-agents porte sur les systèmes intelligents d'extraction d'informations. En effet, les sites Web (sources d'informations) sont généralement passifs en délivrant les pages spécifiques lorsque celles-ci sont demandées par l'internaute. Ces recherches portent donc sur la création de ressources d'information plus « intelligentes » en incluant les capacités des agents. La structure de tels systèmes est décrite dans la figure 1.6.

Figure 1.6: Architecture typique d'un système d'information multi-agent (adapté de Wooldridge, 2002)



Dans la figure 1.6, les répertoires correspondent à des sites Web ou des bases de données. L'accès à ces répertoires est effectué par des agents d'information. Ces agents, qui communiquent grâce à un langage de communication, sont des experts de leur répertoire. Ils sont capables aussi de répondre à des requêtes sophistiquées sur le contenu de leur répertoire (de type: connaissez-vous 'X' ?). Pour trouver des agents dans l'environnement du Web,

certains chercheurs (Wiederhold, 1992; Kuokka and Harada, 1996) proposent des agents intermédiaires ou des courtiers. Chaque agent annonce ses capacités aux courtiers qui sont de différents types : soit ils apparient simplement les annonces de requêtes aux annonces de capacités des courtiers, soit ils collectent simplement de l'information et font des requêtes, soit ils effectuent les deux en même temps. Les courtiers peuvent être spécialisés dans certains domaines d'expertise. Comme le montre la figure 1.6, l'utilisateur effectue une requête sur son ordinateur local. L'agent peut soit contacter directement les agents d'information, soit faire appel à un courtier. Celui-ci va contacter les agents d'information pour savoir s'ils possèdent les compétences requises et fournir ensuite les réponses adéquates. D'autres recherches dans le domaine des SMA portent sur les agents au sein du commerce électronique. Ces recherches sont fondées sur l'idée de développer Internet avec un objectif commercial, ce qui n'était pas son but au départ. Ceci a entraîné la création de systèmes de commerce électronique permettant de commercialiser des produits sur des sites Web. La première génération de ces systèmes multi-agents concerne les systèmes permettant aux internautes d'accéder à des catalogues de produits, de sélectionner un produit, de les payer avec une carte de crédit. La deuxième génération de systèmes consiste à automatiser certains éléments du comportement d'achat du consommateur. Guttman et al. (1998) montrent que plusieurs phases dans le processus d'achat peuvent être automatisés par des SMA. Le tableau 1.10 synthétise leur approche.

Tableau 1.10: Agents de commerce électronique (adapté de Guttman et al., 1998)

	Logique personnelle	Luciole	Agents de marchés	Jango	Kasbah	Agents d'enchères	Tête-à-tête
Besoin d'identification	x	x		x			x
Recherche d'informations produit	x			x			x
Recherche de vendeur			x	x	x	x	x
Negociation					x	x	x
Achat et livraison							
Service et evaluation							

Comme le montre le tableau 1.10, les agents de commerce électronique peuvent intervenir dans la phase d'identification du besoin, dans la recherche d'informations sur les produits, dans la phase de recherche d'un vendeur, pour aider l'internaute à négocier, pour faciliter l'achat et la livraison, et enfin pour le service et l'évaluation. Noriega et Sierra (1999) montrent que ces agents aident le consommateur à automatiser ces différentes phases et à obtenir la meilleure transaction possible. Un des types d'agents simple est l'agent de comparaison d'achat. Supposons qu'un internaute souhaite acheter un CD de Madonna,

l'agent de comparaison d'achat va chercher plusieurs boutiques en ligne et trouver la meilleure offre possible. Ces agents fonctionnent correctement lorsque la comparaison des produits s'effectue sur un seul attribut (comme le prix), mais sont moins performants lorsque plusieurs attributs sont sélectionnés. Doorenbos et al. (1997) proposent le système Jango, système d'agent de commerce électronique de première génération, dont les objectifs sont les suivants:

- aider l'utilisateur à décider quel produit acheter.
- trouver les spécifications des produits
- faire des recommandations à l'utilisateur
- comparer les offres des vendeurs
- analyser les offres spéciales et promotionnelles.

Dans les systèmes d'agents de commerce électronique de deuxième génération, les agents sont capables d'assister la phase de négociation. Le consommateur délègue à l'agent la négociation de la transaction. Cet agent négocie avec un autre agent de manière à obtenir un accord. Les auteurs montrent que ces systèmes ne fonctionnent que si les consommateurs ont confiance dans les agents.

D'autres chercheurs ont développé des agents d'enchères, pouvant participer à des enchères en ligne sur des produits. Un exemple de ces agents est Kasbah, développé par Chavez et Maes (1996). L'objectif de Kasbah est de développer un système dans lequel les utilisateurs créent des agents pour acheter et vendre à leur place. Dans Kasbah, l'utilisateur peut paramétrer les trois éléments suivants pour les agents vendeurs : date de vente, prix souhaité, prix minimum de vente. L'agent commence alors à proposer un produit aux enchères au prix désiré, puis le prix st systématiquement réduit au prix minimum souhaité par le vendeur. L'utilisateur peut spécifier la courbe de baisse du prix offert (linéaire, quadratique ou cubique) et confirmer la vente, grâce à un dernier droit de veto sur le comportement de l'agent. De la même manière, plusieurs paramètres peuvent être fixés pour les agents acheteurs : date d'achat, prix souhaité, prix maximum. De même, l'acheteur peut fixer sa courbe de croissance des prix dans le temps. Les agents de Kasbah agissent dans une place de marché qui gère les enchères en cours. Lorsqu'un acheteur ou un vendeur rentrent dans la place, Kasbah apparie les demandes de produits à acheter avec celles des produits à vendre et fait communiquer les vendeurs et acheteurs correspondants.

Un autre courant de recherche sur l'application des SMA dans un environnement Internet porte sur le développement d'agents d'interfaces homme-machine. Ces agents transforment les systèmes informatiques en assistants proactifs. L'objectif est d'avoir des programmes

informatiques qui prennent des initiatives, plutôt que d'attendre que l'utilisateur précise sa demande. Ces programmes coopèrent donc avec l'utilisateur pour accomplir une tâche. Ces agents sont aussi nommé assistants experts. Maes (1994) définit ces agents par la métaphore de l'assistant personnel qui collabore avec l'utilisateur dans le même environnement de travail. Negroponte (1995), directeur du laboratoire Media du MIT, fut un acteur clef de ce courant de recherche sur les agents d'interface.

Certains chercheurs ont développé également des agents pour des environnements virtuels. Ce courant est lié à la croissance des jeux vidéo et de la réalité virtuelle. L'idée est de créer des agents crédibles, donnant une impression de vie réelle. Une composante essentielle de ces agents est l'émotion : ils doivent montrer des émotions, agir et réagir selon notre empathie envers eux, comprendre les comportements humains. Bates et al. (1992) ont développé plusieurs architectures pour l'émotion (projet OZ par exemple) et crée plusieurs prototypes.

Enfin, plusieurs recherches sur les SMA portent sur des agents dans le cadre de simulations sociales. L'idée est de créer des systèmes multi-agents qui peuvent être utilisés pour simuler le comportement de sociétés humaines. Les agents représentent soit des individus soit des organisations. Conte et Gilbert (1995) montrent que les bénéfices d'une simulation multi-agents des processus sociaux sont les suivants :

- on peut trouver des alternatives à des phénomènes observés dans la nature,
- des propriétés difficiles à observer dans la nature peuvent être étudiées de manière isolée, avec des nouvelles simulations si nécessaire,
- la « socialité » peut être modélisée explicitement : les agents peuvent avoir des représentations d'autres agents. On peut donc étudier les implications de cette représentation.

Un exemple de système de simulation sociale est celui du projet EOS développé par Doran (Doran, 1987; Doran et al., 1992; Doran et Palmer, 1995). L'objectif de ce projet est d'étudier les causes de l'émergence de la complexité sociale au paléolithique en France il y a 15 000 et 30 000 années. A cette époque, on constate en effet une croissance de la complexité sociale des sociétés. La question posée par les archéologues porte sur les causes de cette complexité. Pour répondre à cette question, une plate-forme expérimentale multi-agents (EOS) a été développée. Elle permet de programmer des agents sur la base d'un système de règles. Chaque agent a une représentation symbolique de son environnement: les croyances sur les autres agents (modèle social), les croyances sur les ressources de l'environnement (modèle de ressources), et les autres ressources. Pour actualiser ces croyances, les agents ont des règles cognitives qui transforment les anciennes croyances en nouvelles croyances. Les règles

cognitives et les règles d'action sont exécutées de manière successive. Les agents sont situés dans un environnement à deux dimensions. Les agents occupent une cellule, initialement attribuée au hasard. Ils possèdent des compétences. L'idée est qu'un agent va obtenir des ressources (nourriture) qui sont situées dans plusieurs environnements, et que seuls les agents d'un certain type obtiendront certains types de ressources. Certaines formes de ressources sont associées à un profil de compétences. Ce profil définit pour chaque type de compétence combien d'agents ayant cette compétence sont requis pour obtenir la ressource. Par exemple, la ressource « poisson » exige deux compétences (capacité d'avoir deux bateaux), la ressource chevreuil exige une compétence (avoir une lance). Les agents peuvent donc collaborer pour obtenir les ressources selon des règles prédéfinies.

Un nombre important de phénomènes sociaux ont ainsi pu être analysés par EOS.

Les systèmes multi-agents font donc l'objet de nombreuses recherches touchant des problématiques diverses. Mandiau, Grislin-Le Strugeon et Péninou (2002) distinguent plusieurs grands courants que nous allons présenter.

Les SMA sont tout d'abord utilisés pour modéliser des organisations humaines complexes. Le modèle « holonique » constitue un courant de recherche actif dans ce domaine et définit tout système complexe en tant que hiérarchie flexible, c'est à dire en tant que structure pyramidale au sein de la quelle les membres possèdent des niveaux de responsabilités et une relative autonomie. Les règles organisationnelles définissant ce type de système peuvent être intégrées dans les organisations multi-agents. Le SMA possède alors une structure proche de l'organisation humaine. D'autres chercheurs ont étudié la coopération entre agents via un modèle issu de la théorie des jeux. Ces chercheurs étudient en particulier les agents et les dilemmes sociaux. Un autre courant porte sur la coordination des actions et la planification multi-agents. Dans ce cas, les agents génèrent d'abord leurs plans et cherchent à les coordonner ultérieurement en vue de les exécuter. La cohérence du plan global repose sur les interactions positives et/ou négatives pour anticiper la résolution de conflits potentiels.

D'autres recherches portent encore sur l'implantation de systèmes complexes qui supposent des systèmes fiables et sans erreur. Ces recherches concernent surtout les aspects formels des systèmes multi-agents, comme la spécification, la conception, l'implémentation et la validation de ces systèmes. Certaines recherches étudient le temps dans les SMA, les problèmes de satisfaction de contrainte (planification, conception d'emploi du temps, conception d'interfaces, …). D'autres courants enfin étudient le routage des agents dans les réseaux de communication (notamment sur Internet), la simulation de trafic automobile, la conception de sites robotisés ou l'apport des agents mobiles dans le commerce électronique.

La littérature en marketing comprend également tout un courant de recherche dédié aux agents virtuels qui sont des intermédiaires d'achat pour les consommateurs qui achètent sur Internet. Nous ne développerons pas ce courant car il n'est pas directement dans le champ de la thèse. Il en est de même pour les approches nombreuses sur les réseaux, que nous ne présenterons pas par souci de concision. L'ensemble de ces deux courants de recherche peut cependant apporter une vision complémentaire pour analyser et modéliser des systèmes complexes composés d'agents qui interagissent sur Internet.

Après avoir décrit les approches multi-agents, nous allons maintenant présenter une dernière source théorique que nous utiliserons ultérieurement dans cette thèse. Elle regroupe les modèles utilisés dans la littérature en marketing pour décrire le comportement des internautes et l'interactivité internaute-site Web. Certains de ces modèles seront effectivement repris dans les applications de la thèse qui font l'objet des chapitres 3, 4 et 5.

1.7 LES MODELES DE COMPORTEMENT DES INTERNAUTES DANS LA LITTERATURE EN MARKETING:

Durant les dix dernières années, les professionnels du marketing et les chercheurs ont tenté de mieux comprendre les phénomènes liés à Internet pour pouvoir développer des stratégies adaptées à ce nouveau médium de communication et de distribution. Le thème principal de ces recherches est l'amélioration du design des sites dans l'objectif de maximiser les fréquences de visite, la durée des visites et les achats. Nous présenterons dans un premier temps les principaux articles dans ce courant de recherche. Dans un deuxième temps, nous nous concentrerons sur les modèles utilisés en marketing pour étudier le comportement de visites et d'achat sur les sites Web, que nous utiliserons dans les applications empiriques qui constituent les prochains chapitres de la thèse.

1.7.1 Les recherches sur l'amélioration des sites Web

Nous citerons dans ce courant quelques articles principaux. Mandel et Johnson (2002) étudient par exemple l'influence du design d'une page Web sur le choix de produits par le consommateur. Ils mènent plusieurs expériences en manipulant les couleurs et les images de fond de pages Web et montrent l'influence des visuels sur le choix des produits, que ce soit pour les novices ou les experts (pour les novices, les visuels influent la recherche externe et

modèrent les changements de préférence, tandis que pour les experts, la recherche externe reste inchangée mais les visuels modifient le choix des produits).

Gorn, Chattopadhyay, Sengupta et Tripathi (2004) analysent le lien entre la couleur de fond de la page Web et la rapidité perçue du téléchargement de cette page. Ils montrent d'une part l'existence de liens entre la couleur et les sensations de relaxation puis entre la sensation de relaxation et la perception du temps. Les auteurs prédisent ainsi que la couleur du fond de la page influence la perception de rapidité de téléchargement, la sensation de relaxation modérant cette influence. Gorn et al. (2004) montrent également que la couleur a des conséquences sur l'évaluation d'un site Web et sur la probabilité de recommander le site à d'autres personnes.

Drèze et Zufryden (2004) proposent même une mesure de la visibilité on-line et étudient son impact sur le trafic d'un site. Ils définissent la visibilité comme le degré auquel l'utilisateur est prêt à considérer une référence de site Web dans son environnement on-line et off-line. L'index de visibilité qu'ils proposent dépend de plusieurs éléments : les bannières de publicité sur Internet ; les résultats de recherche du site sur un moteur de recherche, de la liste fournie dans le moteur de recherche, des liens vers le site depuis d'autres sites, des groupes de discussions et des forums, des e-mails reçus, des articles on-line. Ils montrent que cette visibilité on-line est très liée au trafic d'un site Web.

Un article très souvent cité dans les recherches sur le Web est celui d'Ansari et Mela (2003) qui étudient la personnalisation du contenu des e-mails et sa capacité à accroître le trafic des sites Web. Ces auteurs utilisent des données de navigation des internautes et développent un modèle probit MDP (adapté du modèle de Dirichlet) couplé à une approche d'optimisation pour personnaliser l'information sur Internet. Ils montrent notamment que la customisation de l'information des e-mails augmente le nombre de clicks de 62%.

D'autres auteurs, comme Johnson, Bellman et Lohse (2003), s'intéressent à l'effet de l'apprentissage sur les sites Web. Ils montrent que la plupart des sites Web se caractérisent par une baisse des temps de visite et que les sites caractérisés par une forte courbe d'apprentissage sont ceux qui ont les plus forts taux d'achat.

Toujours dans ce même courant de recherche sur l'amélioration des sites Web se situent plusieurs articles s'intéressant à la recherche d'informations. Lynch et Ariely (2000), Zauberman (2003) mesurent l'impact des coûts de recherche d'informations sur Internet. Ce courant repose sur l'idée selon laquelle l'un des avantages majeurs d'Internet est la réduction du coût de recherche d'informations pour les consommateurs. Utilisant une expérimentation sur le marché des vins on-line, Lynch et Ariely (2003) font varier indépendamment trois

types de coûts: le coût de recherche de l'information prix, le coût de recherche de l'information sur la qualité et le coût de recherche pour comparer deux sites Web de vente de vins. Les auteurs en déduisent quatre résultats principaux : sur le marché des vins, baisser le coût de recherche d'information sur la qualité réduit la sensibilité au prix ; la sensibilité au prix sur les vins communs aux deux magasins augmente lorsque la comparaison d'informations est facilitée ; créer des environnements plus transparents en baissant les coûts de recherche d'information des trois critères produit des gains pour les consommateurs (ils apprécient plus leur expérience d'achat) ; lorsque la comparaison des magasins on-line est difficile, les parts de marché des deux magasins sont proportionnelles à leur part de distribution, mais lorsque la comparaison est facilitée, la part de marché de l'un des magasins décroît significativement. Zauberman (2003) s'intéresse plutôt aux modifications dynamiques dans la structure de coût d'information (trade-off des coûts initiaux d'entrée et des coûts d'utilisation) et à la manière dont les préférences en fonction du temps affectent le comportement de recherche d'information et de switch des consommateurs. Ils définissent la propension du consommateur à rechercher des informations et à switcher de site par le terme «mise sous clef» (traduction que nous avons adoptée pour le terme anglais « lock-in »). Zauberman (2003) montre au travers de ses trois expériences que les consommateurs tendent à minimiser leurs coûts présents et n'incorporent pas les coûts futurs de switch dans leurs décisions. Le principal message de cette recherche est que les préférences inter-temporelles dynamiques constituent un déterminant du comportement des consommateurs. Ceux-ci ne réalisent pas qu'ils seront affectés dans le futur par des investissements actuels liés à une alternative.

Bakos (1997) étudie les systèmes d'information intermédiaires entre l'acheteur et le vendeur qui créent une place de marché électronique diminuant les coûts de recherche d'informations sur les prix du vendeur et l'offre produit. Il montre que ces places de marché électroniques réduisent les inefficacités dues aux recherches d'informations de l'acheteur, réduisent la capacité des vendeurs à faire des profits monopolistiques et accroissent la capacité des marchés à allouer les ressources optimales et à être plus efficaces.

Wu et al. (2004) étudient le problème de "free riding" sur Internet. Selon eux, fournir de l'information sur Internet coûte cher aux entreprises. Un consommateur peut pourtant être convaincu d'acheter un produit grâce aux informations fournies par un site de vente, mais finalement acheter le produit sur un autre site où le prix est moins cher. Les auteurs proposent un modèle de marché où l'information est coûteuse et possède les caractéristiques d'un bien public. Les consommateurs utilisent le service d'information pour identifier le produit idéal,

mais peuvent ensuite chercher un prix moins élevé grâce à Internet, la recherche d'information étant réduite. Les auteurs montrent que dans ce cadre, un vendeur doit s'identifier comme fournisseur d'information s'il veut faire des profits. D'autre part, lorsqu'il y a une plus grande compétition sur le marché de l'information fournie, les vendeurs ont une moindre motivation à fournir de l'information.

Les quelques recherches que nous avons décrites ci-dessus se concentrent sur une problématique précise du Web (essentiellement comment améliorer le design et le contenu d'un site Web) mais ne cherchent pas à représenter ou à modéliser le comportement de l'internaute dans son ensemble et dans le système global qu'est Internet.

Adoptant donc une approche globale, Novak, Hoffman et Yung (2000) développent un modèle de ce type en utilisant des équations structurelles pour mesurer l'expérience du consommateur dans des environnements on-line. Leur modèle s'appuie sur la théorie du flux que ces auteurs avaient développée dans des recherches précédentes et que nous avons précédemment présentée. Dans cette recherche, Novak, Hoffman et Yung (2000) analysent de manière plus approfondie le concept du flux pour le modéliser. Le flux sur le Web est selon eux un état cognitif expérimenté pendant la navigation et qui dépend de quatre éléments : le niveau des compétences et du contrôle ; le niveau des défis et de l'attention ; l'attention focalisée ; l'interactivité et la télé présence. Ils montrent que la concentration sur la navigation, qui permet l'expérience de flux, est fondamentale sur Internet : les pensées ou les perceptions du consommateur qui ne relèvent pas de la navigation sont écartées et le consommateur focalise son attention sur l'interaction. Leur modèle conceptuel montre le degré auquel l'expérience de navigation peut être définie, mesurée et reliée à d'autres variables marketing.

1.7.2 Les modèles de comportement de visite et d'achat sur le Web

De nombreuses recherches récentes en Marketing portent sur l'analyse du comportement de visite et d'achat sur le Web. Ce thème est important pour plusieurs raisons, à la fois sur le plan managérial et académique. Tout d'abord, sur le plan managérial, les entreprises cherchent aujourd'hui à utiliser le Web comme un nouveau médium de distribution ou de communication, mais certaines d'entre elles hésitent à aller plus loin en raison de leur manque de connaissance du comportement de l'internaute et de ses effets potentiels. Ces entreprises ont pourtant intérêt à modéliser les visites on-line pour mesurer les effets de leur communication sur le Web et l'impact financier de leurs actions sur Internet. Face à ce

manque de connaissances, plusieurs chercheurs ont essayé de mieux comprendre le comportement des internautes. Ces recherches ont développé notamment la capacité à suivre le comportement des internautes, en particulier en ce qui concerne les visites et l'achat. Nous pouvons noter que le courant de recherches sur le comportement de visites et d'achats est très actif en marketing, puisque la plupart des articles détaillés ci-dessous ont été publiés en 2003 et 2004, dans les meilleures revues internationales (revues majoritairement de rang alpha). Nous constatons également que ces recherches concernent un nombre limité de chercheurs qui ont publié durant ces deux années plusieurs articles sur ce thème dans les meilleures revues (comme Fader ou Moe par exemple).

Nous avons synthétisé ces recherches dans le tableau 1.11 ci-dessous :

Tableau 1.11: Synthèse des modèles utilisés pour étudier les comportements de visites et d'achat

Auteurs	Année de publication	Journal	Thème	Un seul site/ Multisites	Variables modélisées	Type de modèle
Bucklin et Sismeiro	2003	JMR	Navigation	Un seul site	Requête de page/ Probabilité de sortie du site Durée des pages vues	Tobit type II
Telang, Boatwright et Mukhopadhyay	2004	JMR	Visites	Un seul site	Visites	PHM mixtes
Sismeiro and Bucklin	2004	JMR	Achat	Un seul site	Achats NUT (Nominal Tasks User)	Modèle séquentiel de probits binaires
Moe et Fader	2004	Management Science	Conversion des visites en achats	Un seul site	Effets accumulés des visites Effets de seuil d'achat	Modèles stochastiques dynamiques
Moe et Fader	2004	Journal of Interactive Marketing	Modèle évolutif de comportement des visites (modèle EV)	Un seul site	Comportement de visite Gain d'expérience	Modèle Exponentiel gamma étendu
Montgomery et al.	Forthcoming	MS	Trajet de navigation	Un seul site	Comportement permanent Changements abrupts de navigation	Modèle probit multinomial dynamique (incluant un modèle VAR et des chaînes de Markov)
Park et Fader	2004	MS	Navigation	Multisites	Processus de temps de viste Propension de visite entre deux sites	Approche de Sarmanov
Johnson, Moe, Fader, Bellman et Lohse	2004	Management Science	Comportement de recherche	Multisites	Profondeur de recherche Dynamique de recherche Activité de recherche	Modèles de probabilités

Bucklin and Sismeiro (2003) modélisent deux facteurs clés du comportement de navigation sur un site Web : la décision de l'internaute de continuer la navigation (ou encore la requête de page) et la durée de chaque page vue. Utilisant des données de navigation issues de fichiers logs, les auteurs utilisent dans leur approche un modèle Tobit type II. Ils modélisent tout d'abord la décision de navigation (probabilité de requête d'une nouvelle page ou probabilité de quitter le site) par un modèle probit hiérarchique binaire permettant d'estimer les paramètres propres aux visiteurs. La durée de la page vue par chaque internaute est représentée avec le même type de modèle mais les covariés sont liés à des caractéristiques de l'internaute et aux spécificités de la page vue (comme par exemple le temps requis pour la transmission de données, l'activité de l'utilisateur, l'accomplissement des tâches). Bucklin and Sismeiro (2003) montrent d'une part l'effet d'apprentissage lié à la répétition des visites. D'autre part, l'effet de la visite de répétition n'est pas le même sur les deux composantes du comportement de navigation : lorsque les visiteurs reviennent sur le site, ils réduisent le nombre de pages vues mais pas la durée de chaque page vue. Ils notent également que les visites répétées ont un impact négatif sur la probabilité de requête d'une page additionnelle. Enfin, les tâches de chaque page prennent le même temps au visiteur malgré ses visites précédentes. Bucklin and Sismeiro (2003) confirment également la théorie de Zauberman (2003) du « lock-in » de l'internaute à l'intérieur d'un site (ci-dessus présentée). Ils montrent en effet que la profondeur de la visite (nombre cumulé de pages vues pour arriver à la page donnée) a un impact positif sur la durée moyenne de visite : lorsque les visiteurs font des requêtes de plusieurs pages, ils ont tendance à passer plus de temps sur chaque page. La profondeur de la visite a également un impact négatif sur les requêtes de pages additionnelles dans le site. Leur recherche révèle aussi des caractéristiques de navigation qui reflètent des stratégies de gain de temps de l'internaute : plus les utilisateurs restent sur le site, plus la probabilité de sortir du site (à partir d'une page donnée) augmente.

D'autres auteurs ont analysé le comportement de visites sur les sites Web. Telang, Boatwright, and Mukhopadhyay (2004) étendent la littérature en marketing en développant des modèles proportionnels de hasard (PHM) mixtes. Ces modèles permettent de capturer sur les achats des périodicités et une hétérogénéité non observée et sont appliqués pour prédire les visites sur les moteurs de recherche en incluant le planning horaire de l'utilisateur. Relativement aux approches traditionnelles, le modèle propose par Telang, Boatwright, and Mukhopadhyay (2004) apportent trois contributions principales : d'une part, ils fournissent un modèle plus complet pour prédire les visites de moteurs de recherche ; d'autre part le modèle

mixte tient compte de pics de données et permet de prédire des données désagrégées de manière plus précise ; enfin, les covariés ajoutent de la précision dans les prédictions. Les auteurs dévoilent également plusieurs caractéristiques des visites : les consommateurs ont tendance à diminuer leur utilisation de moteurs de recherche avec le temps. D'autre part, la pratique de tâches de non-recherche (comme l'e-mail, le chat, les informations, les cotations d'actions) accroît la fréquence de visites et l'utilisation régulière du site, les individus effectuant ces actions n'ayant pas tendance à moins utiliser les moteurs de recherche avec le temps.

Park et Fader (2004) s'intéressent à une problématique différente : le comportement de visite non pas sur un seul mais sur différents sites Web. Ils cherchent à analyser notamment en quoi l'information sur un site explique le comportement de visite sur un autre site. Leur modèle stochastique de temps capture deux sources d'association dans les critères de navigation: l'observation de processus dans le temps (comme les temps d'arrivée sur le site), les propensions de visites (variable latente) sur différents sites concurrents. Le modèle proposé est une généralisation du modèle univarié exponentiel-gamma. Les auteurs montrent que la non prise en compte de ces deux sources d'association entraîne de mauvaises prédictions de visites et génère des paramètres biaisés. Le modèle de Park et Fader (2004) se fonde sur les distributions multivariées de la famille de Sarmanov, plus flexibles que les autres distributions bivariées (comme les distributions de la famille Farlie-Gumber-Morgenstern (FGM) utilisées fréquemment dans les modèles bivariés). Le modèle proposé utilise l'approche de Sarmanov pour construire chacune des deux composantes du modèle : le processus de temps de visite au travers des différents sites (au niveau individuel) et la distribution de l'hétérogénéité des propensions de visite entre deux sites sur l'ensemble des individus. Grâce à ce modèle, les auteurs peuvent ainsi prédire les visites futures et corriger la surestimation du nombre de visites souvent faite par les chercheurs qui considèrent comme indépendants les temps de visite sur plusieurs sites. Le modèle de Park et Fader (2004) permet également de mesurer l'acquisition de consommateurs. En effet, les auteurs peuvent prédire l'existence et le comportement d'une classe « zéro » (visiteurs qui n'étaient pas précédemment venu sur le site) à partir de deux variables d'information (nouveauté et fréquence) sur les visites précédentes d'un site concurrent.

Dans le même courant de recherche sur la navigation multi-sites, Johnson, Moe, Fader, Bellman, and Lohse (2004) cherchent à modéliser le comportement de recherche sur plusieurs sites concurrents de e-commerce. Utilisant des modèles de probabilités, ils caractérisent le comportement de recherche au niveau individuel en termes de profondeur de recherche, de

dynamique de recherche et d'activité de recherche. La tendance des individus à faire des recherches est modélisée par une fonction logarithmique. Les auteurs montrent que les acheteurs font des recherches sur très peu de sites et que la propension de recherche est stable dans le temps. Les auteurs modélisent aussi l'activité d'achat d'un foyer et montrent que les acheteurs les plus actifs on line ont tendance à faire des recherches sur plus de sites. Les caractéristiques du consommateur expliquent en grande partie l'aspect dynamique de la recherche.

Sismeiro and Bucklin (2004) proposent un modèle de comportement d'achat sur un site Web de vente de voitures. Utilisant des fichiers log de navigation, les auteurs modélisent l'achat on line en liant la décision d'achat aux actions du visiteur et aux éléments auxquels celui-ci est exposé sur le site. Le processus d'achat est décomposé en trois tâches principales (concept de NUT : Nominal Task User) effectuées séquentiellement par l'internaute : la configuration du produit, la fourniture complète d'informations personnelles, la confirmation d'ordre avec vérification des données de carte de crédit. Utilisant une séquence de probits binaires et des approches Bayésiennes, les auteurs modélisent pour chaque visiteur la décision d'accomplir chaque tâche pour la première fois, sachant que le visiteur a accompli les tâches précédentes au moins une fois. Pour mieux prédire l'achat et comprendre le comportement des visiteurs, le modèle incorpore des covariés sur les éléments auxquels sont exposés les visiteurs lors de la navigation et sur ce qu'ils font lorsqu'ils accomplissent chaque tâche (comportement de navigation, répétition de visite, utilisation d'aides à la décision, effort pour rentrer des données, collecte d'informations et processus pour cette collecte). Sismeiro and Bucklin (2004) incluent également l'hétérogénéité régionale des visiteurs dans leur modèle. Les résultats montrent que les expériences de navigation et le comportement prédisent l'accomplissement des tâches à tous les niveaux de décision. Le nombre de visites n'apparaît pas un bon indicateur pour prédire la propension à acheter et un site proposant des systèmes d'aide à la décision sophistiqués ne garantit pas des taux de conversion plus forts. Les auteurs montrent également que l'approche séquentielle fournit de meilleures prévisions que des modèles non séquentiels. Cette recherche dévoile que des variations de comportement dans l'accomplissement des tâches sur un site Web s'explique par ce que font les visiteurs durant la navigation et par les caractéristiques des pages auxquels ils sont exposés. Les relations entre l'accomplissement des tâches et l'utilisation du site et l'exposition aux caractéristiques des pages constituent des outils de diagnostic pour l'évaluation de la performance du site.

Moe et Fader (2004b) étudient la conversion des visites sur un site en achats. Ils proposent un modèle de comportement de conversion en liant visites et achats dans un modèle dynamique

permettant de prédire la fréquence d'achat et le volume sur la base des visites précédentes et des achats sur le site. Le comportement de conversion est décomposé deux éléments : le premier concerne les effets accumulés des visites, le deuxième les effets de seuil d'achat, chaque élément variant dans le temps et selon les consommateurs. Les effets des visites incluent la notion selon laquelle les visites de magasin jouent différents rôles dans le processus d'achat. Le modèle inclut notamment la différence entre des types de visites, décrite dans des recherches précédentes de Moe (2003). Moe (2003) distingue en effet quatre types de visites d'achat on-line:

- les visites avec un but direct d'achat: l'objectif est d'acheter rapidement,
- les visite de recherche et de délibération : les consommateurs cherchent un produit avec une éventuelle possibilité d'acheter,
- les visites liées à une navigation à caractère hédoniste : l'internaute achète pour le plaisir,
- Les visites pour se construire des connaissances : l'objectif est d'apprendre et de recueillir des informations sur le marché.

Moe et Fader (2004b) incluent donc dans leur modèle le fait que certaines visites ont un objectif planifié d'achat alors que d'autres sont liées à une navigation purement hédonique. Les effets de seuil capturent la résistance psychologique à l'achat on-line qui peut augmenter ou décroître selon l'expérience acquise par le consommateur lors du processus d'achat sur un site donné. Six composantes clés constituent le modèle de conversion de Moe et Fader (2004b) : la probabilité de base d'acheter pour chaque visite ; l'effet positif de la visite sur l'achat (d'une part, chaque visite a un impact stochastique sur l'achat, d'autre part, les effets des visites s'accumulent dans le temps et augmentent la probabilité d'achat) ; l'hétérogénéité dans les effets des visites et les seuils d'achat, qui varient selon les consommateurs ; l'évolution des effets dans le temps (selon les gains d'expérience des consommateurs) ; les consommateurs qui n'achètent jamais (segment d'internautes qui utilisent le site uniquement pour les informations). En incorporant l'hétérogénéité entre les consommateurs et les aspects dynamiques dans le temps, Moe et Fader (2004b) montrent que leur modèle de conversion est plus performant et facilement utilisable que les modèles classiques comme les régressions logistiques. A ce propos, Moe et Fader (2004b), comme de nombreux autres auteurs, soulignent la difficulté de modéliser les comportements sur le Web en raison des bases de données très larges et détaillées. Ils soulignent les qualités de parcimonie et d'efficacité que doivent développer les chercheurs pour mener à bien des modèles statistiques complexes appliqués à l'environnement du Web.

Dans un autre article, Moe et Fader (2004a) proposent un modèle évolutif de comportement de visite d'un site Web. Utilisant des fichiers log de navigation, leur modèle capture les variations dans le comportement de visite du site et le gain d'expérience du visiteur. Ils proposent un modèle non stationnaire, nommé EV (Evolving Visit) basé sur trois composantes : un processus de temps capturant les taux individuels de visites (temps et fréquence) ; la distribution hétérogène qui capte les différences entre les individus ; un processus d'évolution qui permet de modifier les taux de visite par individu d'une visite à l'autre. Le modèle EV proposé est fondé sur un modèle Exponentiel Gamma dans lequel les taux de visites varient dans le temps pour un individu. En terme de résultats, les auteurs montrent d'une part que plus les internautes visitent un site fréquemment, plus leur propension à acheter est importante. D'autre part, l'évolution de la fréquence de visite dans le temps fournit des informations sur les segments des consommateurs qui ont les plus forts taux de conversion d'achat.

Shibo Li (2004) a gagné le prix John A. Howard AMA Doctoral Dissertation Award. Dans sa thèse, l'auteur développe une modélisation du comportement de navigation sur le Web. Un article de Montgomery et al. (2004), issu de ces recherches, sera publié prochainement dans Marketing Science. Utilisant des données de navigation de fichiers log, les auteurs montrent que l'information sur le trajet de la navigation peut être représentée par un modèle probit multinomial dynamique. Deux éléments dynamiques sont inclus dans leur modèle. Tout d'abord, les auteurs introduisent un comportement permanent à travers une composante suivant un modèle VAR. L'idée est qu'une plus grande affinité avec une page Web entraîne plusieurs visualisations de la page. Un utilisateur peut visualiser plusieurs pages consécutivement, introduisant la nécessité d'introduire des séries temporelles dans le modèle. Le deuxième élément est l'incorporation dans le modèle de changements abrupts de comportement de navigation par l'introduction d'un modèle mixte variant dans le temps. Une chaîne de Markov gouverne les transitions entre les états de ce modèle mixte. Par exemple, durant une session, un utilisateur peut changer d'objectifs et décider de se focaliser sur l'achat, ou alternativement de juste regarder ailleurs au lieu d'acheter. Les auteurs estiment le modèle sur les données d'un site de ventes de livres (barnesandnomble.com, données issues du panel d'utilisateurs CMM de Jupiter Media Metrix). Ils montrent notamment que la composante de mémoire du modèle est cruciale pour prédire le chemin de navigation. Leur modèle fournit de meilleurs résultats que des modèles traditionnels multinomiaux probit et des modèles de Markov de premier ordre. Ces résultats suggèrent que le chemin suivi par l'utilisateur permet de prédire les mouvements futurs sur le site. L'une des applications

potentielles est la prédiction du taux de conversion. Montgomery et al. (2004) montrent qu'après six visualisations, le taux de conversion des internautes peut être prédit avec plus de précision que s'il n'y avait aucune information sur le trajet de navigation.

Les modèles cités ci-dessus présentent plusieurs limites, souvent citées dans les articles eux-mêmes. Tout d'abord, la plupart de ces modèles sont complexes. D'autre part, la taille importante des bases de données disponibles sur le Web (Moe et Fader, 2004b) rend leur utilisation difficile et leur implantation en entreprise peu probable. De plus, la plupart des recherches ci-dessus soulignent dans leurs limites la non-prise en compte de tous les éléments de navigation de l'internaute, cet élément étant probablement lié au fait que cette tâche est particulièrement complexe. Face à ces constats, nous allons dans cette thèse introduire les modèles des réseaux de Petri, qui permettent de modéliser des systèmes complexes tout en présentant une représentation graphique simple du phénomène étudié. Dès 1995, Ferber présente les réseaux de Petri comme l'une des techniques de modélisation des systèmes multi-agents (les recherches menées dans ce domaine seront présentées dans le chapitre 2 de la thèse consacré à la présentation des réseaux de Petri). De plus, les réseaux de Petri sont des outils de simulation qui répondent au besoin croissant de modéliser le comportement sur le Web en fonction de différents stimuli marketing.

CONCLUSION :

Nous avons présenté dans ce chapitre des courants de recherches provenant de domaines très divers. Après avoir abordé les théories sur l'interactivité en marketing, nous avons développé les approches en psychologie sociale étudiant les interactions entre des agents au sein de communautés virtuelles. Nous avons ensuite introduit la théorie de la coproduction de valeur fondée sur les interactions entre agents ou entités développée en stratégie, avant de détailler les recherches en informatique et en Intelligence Artificielle sur les systèmes multi-agents, dont l'objectif est d'étudier les interactions entre plusieurs agents d'un même environnement. Enfin, nous avons détaillé dans le domaine du marketing les modèles de comportement des internautes, en particulier les modèles de comportement de navigation, de visite et d'achat qui font l'objet d'un courant de recherche très actif en marketing ces dernières années.

L'approche des réseaux de Petri que nous allons présenter dans le chapitre suivant de la thèse repose sur l'intégration de ces différents courants de littérature et sur la modélisation de systèmes multi-agents interactifs sur le Web dans la perspective d'un système global.

Le fait de modéliser le comportement des internautes dans la perspective d'un système global est intéressant pour plusieurs raisons. Tout d'abord, le comportement de visites et d'achats sur le Web diffère du même comportement dans les magasins réels, en particulier sur le critère du taux de conversion des visites en achat, faible sur le Web. En effet, des millions de visiteurs visitent des sites Web, mais 3% uniquement achètent (selon une étude du Boston Consulting Group et shop.org (Betts, 2001)). Ceci implique le fait que, comme le soulignent Sismeiro and Bucklin (2004), les visites sur le Web ne sont pas de bons indicateurs pour prédire des intentions d'achat, et que les modèles single-stage (e.g., probit, logit) mesurant un impact direct des visites sur le comportement d'achat ne sont pas adaptés pour analyser les comportements sur le Web. Une deuxième différence entre les comportements on- and off-line concerne l'objectif de la visite. Moe (2003) distingue quatre types de visites d'achat on-line:

- les visites avec un but direct d'achat: l'objectif est d'acheter rapidement,
- les visite de recherche et de délibération : les consommateurs herchent un produit avec une éventuelle possibilité d'acheter,
- les visites liées à une navigation à caractère hédoniste: l'internaute achète pour le plaisir,
- Les visites pour se construire des connaissances: l'objectif est d'apprendre et de recueillir des informations sur le marché.

Le comportement de visites dans un magasin “brick-and-mortar” est différent de celui des internautes, généralement plus concentré sur un achat direct: les consommateurs visitent rarement un supermarché dans un but purement exploratoire. La recherche de Moe (2003) montre que chaque visite sur le Web a un but spécifique et que le taux de conversion d’achat doit dépendre du type de visite effectuée.

Les caractéristiques spécifiques du comportement de visite et d’achat sur le Web nécessitent donc des méthodes flexibles et adaptées pour modéliser les actions des différents acteurs du système considéré. La première caractéristique est la simultanéité des actions. En raison de l’espace géographique illimité du Web, plusieurs actions se produisent en même temps. Par exemple, un nombre quasi illimité de visiteurs peuvent naviguer en même temps sur la page Web d’un produit, demander des informations supplémentaires à un agent intermédiaire du Web, acheter le produit simultanément. Pour toutes ces actions parallèles, les systèmes Web doivent pouvoir intégrer plusieurs visites parallèles qui doivent être incorporées dans un processus de séquences différentes et parfois synchronisées, comme par exemple la vérification des informations sur les cartes de paiement à convertir en achat. Par conséquent, les comportements de visites et d’achat sur le Web reflètent des systèmes qui sont souvent trop complexes à modéliser de manière analytique parce qu’ils requièrent des ressources informatiques trop importantes. Les modèles classiques de causalité, en particulier les modèles single-stage, qui explorent une partie d’un système complexe, sont limités pour analyser l’aspect dynamique des comportements de visites et d’achat sur le Web. Pour répondre à ces problématiques, nous proposons dans cette thèse d’explorer les modèles des réseaux de Petri qui peuvent analyser des systèmes complexes, en particulier s’ils sont constitués d’actions parallèles, séquentielles, ou encore synchronisées. Nous allons donc consacrer le chapitre 2 de la thèse à exposer les modèles des réseaux de Petri et à présenter une revue de littérature sur ce type d’approche.

Chapitre 2 :
Les réseaux de Petri : modèle et revue de littérature

1.1 INTRODUCTION AUX RESEAUX DE PETRI:

Les réseaux de Petri furent créés dans les années soixante par Carl Adam Petri (1962) pour étudier des systèmes dynamiques complexes. Ils furent ensuite développés au MIT et appliqués dans de nombreux domaines. Cette technique de modélisation est aujourd'hui l'une des principales utilisées pour des analyses à la fois qualitatives et quantitatives de processus dans lesquels des informations, des documents ou des tâches sont transmises d'un individu à un autre, ou d'une machine à une autre, l'ensemble du système évoluant vers une nouvelle phase à chaque étape. Parmi l'ensemble des modèles existants, les réseaux de Petri et leurs extensions possèdent un intérêt fondamental indéniable, car, selon Diaz (« Les réseaux de Petri, modèles fondamentaux », 2001):

- ils ont fourni toutes les premières approches de modélisation, ainsi que leurs sémantiques, utilisées pour modéliser et maîtriser les comportements des systèmes parallèles et distribués, synchronisés et communicants ;
- ils définissent un support graphique aisé pour l'expression et la compréhension des mécanismes de base de ces comportements ;
- ils constituent des représentations faciles à comprendre et donc à manipuler, à la fois pour la création des modèles et pour leur analyse ;
- ils permettent d'exprimer très simplement les concepts premiers des fonctionnements communicants, en incluant les phénomènes d'attente et de synchronisation, et en prenant en considération leurs caractéristiques et paramètres temporels et stochastiques ;
- étant non liés à un langage particulier de réalisation, ils assurent l'indépendance de la modélisation vis-à-vis des applications;
- ils possèdent, pour les étapes de conception, des méthodes de validation basées sur un grand nombre de résultats théoriques et d'outils de support, utilisables pour étudier les comportements fonctionnels, temporels et stochastiques ;
- ils possèdent un grand intérêt dans de très nombreux et importants domaines d'applications.

Les réseaux de Petri ont été utilisés dans de nombreuses recherches, particulièrement dans les domaines suivants : recherche opérationnelle, informatique, systèmes d'information mais aussi science des organisations. Un site Web (www.daimi.au.dk/PetriNets) particulièrement complet sur les réseaux de Petri regroupe une communauté virtuelle de l'ensemble des chercheurs et praticiens qui dans le monde travaillent sur ce thème.

Les réseaux de Petri peuvent être utilisés pour modéliser et simuler des systèmes de tout type. Ils sont en particulier très utiles pour décrire et analyser des systèmes dynamiques complexes qui comportent des flux d'objets ou d'informations. Il existe également un important formalisme mathématique associé aux réseaux de Petri, sur lequel plusieurs recherches ont été et sont toujours menées actuellement.

Enfin, deux conférences internationales sont entièrement dédiées aux réseaux de Petri : ATPN, Application of the Theory of Petri Nets ; PNPM, Petri nets and Performance Models.

Dans ce chapitre, nous allons dans une première partie décrire le principe des réseaux de Petri et l'illustrer par deux exemples simples. Nous détaillerons ensuite les différentes familles des réseaux de Petri. Nous approfondirons dans une troisième partie la famille des réseaux de Petri stochastiques, que nous utiliserons ultérieurement dans les applications développées dans la thèse. Enfin, nous synthétiserons différents articles issus d'une revue de littérature sur les réseaux de Petri, volontairement limitée aux articles dont les problématiques nous ont semblé intéressantes par rapport au champ du marketing qui constitue notre domaine de recherche.

2.2 PRESENTATION DES RESEAUX DE PETRI :

2.2.1 Eléments de base :

Les réseaux de Petri permettent de modéliser des systèmes dynamiques, c'est à dire des systèmes évoluant d'un état à un autre à la suite d'évènements internes ou externes.

Un réseau de Petri est constitué des composantes structurelles suivantes :

- *Les places*: notées graphiquement par des cercles ; les places représentent les différents états ou conditions du système ;
- *Les transitions*: notées graphiquement sous forme de barres ou de rectangles, les transitions décrivent les évènements qui modifient les états du système ;

- *Les arcs*: notés par des flèches qui joignent les places aux transitions et les transitions aux places, ils constituent les connecteurs entre une place et une transition dans un sens déterminé;
- Une distribution de *jetons* dans les places, dont le rôle est décrit dans la partie 2.2.2 au travers d'un exemple.

2.2.2 Exemple : modélisation d'un mécanisme de wish list sur Internet (modèle de base)

Le concept de wish list :

L'utilisation croissante des wish lists est étroitement liée au développement d'Internet. Le principe d'une wish list est qu'un individu dépose une liste de souhaits de cadeaux sur un site Web. Amazon.fr, par exemple, propose sur son site une rubrique « vos envies cadeaux », cette technique marketing étant utilisée sur plusieurs sites de vente en ligne de produits grand public. Certains sites comme www.Millemercis.com ont pour activité principale la gestion de wish lists. Ils proposent aux internautes de constituer une listes de cadeaux à partir de n'importe quel site marchand. Les internautes ayant déposé une liste chez Millemercis peuvent intégrer à la liste tout produit venant d'un site Web quelconque grâce à un logiciel spécifique appelé « facilitateur ». Grâce à ce système, 500 000 internautes ont constitué une liste de cadeaux chez Millemercis. Cette entreprise possède ainsi une base de données particulièrement intéressante en marketing : on connaît par individu l'ensemble des produits souhaités, toutes catégories de produit confondues (ce type de base de données est relativement unique). Ayant déposé sa liste de cadeaux, l'internaute peut fournir les e-mails de personnes susceptibles de lui faire un cadeau (amis, relations, famille, collègues, etc…). Ces personnes ont alors la possibilité de consulter la wish list afin d' offrir (ou non) le cadeau. Tous les agents intervenant dans ce système en tirent des bénéfices : celui qui a déposé la liste peut recevoir un cadeau, l'acheteur du cadeau a la satisfaction d'offrir un produit souhaité (bénéfices hédoniques) et l'entreprise vend ses produits. De plus, elle commercialise ses produits via des internautes fournissant eux-mêmes une liste d'acheteurs potentiels (puisqu'ils augmentent la probabilité de recevoir un cadeau de leur liste). La wish list est un système multi-agents : un premier agent est l'individu qui dépose la wish list sur le site; son objectif est de recevoir le plus rapidement possible un cadeau. Le deuxième agent est l'entreprise qui propose des produits sur la wish list ; son objectif est de vendre le maximum de produits en un minimum de temps. Enfin, le troisième type d'agent regroupe

tous les amis, relations ou connaissances du premier agent qui viennent consulter la wish list et éventuellement faire un cadeau ; ces agents soit refusent de participer au système, soit ont envie de faire un cadeau au premier agent. L'objectif global d'un système de wish list consiste à maximiser la fluidité des échanges, à éviter les points de blocage (comme par exemple le non-achat d'un produit par un agent ayant consulté la liste), de manière à ce que le maximum de produits soient achetés en un minimum de temps, conciliant ainsi les objectifs des trois types d'agents. On constate ici l'intérêt marketing pour une entreprise de modéliser ce phénomène avec précision et de chercher à l'optimiser.

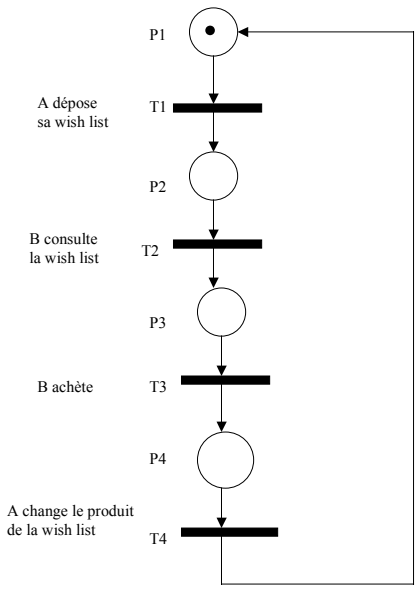
Modélisation de la wish list par un réseau de Petri simple :

Un mécanisme de wish list simple et fonctionnant sans blocage peut être décrit de la manière suivante :

Etat initial du système: le système est composé d'un internaute (noté A), d'un ami de cet internaute (B), d'un produit commercialisé par une entreprise.
Etape 1 : L'internaute A dépose sa wish list (produit P) sur le site
Etape 2 : L'ami B consulte la wish list de A
Etape 3 : L'ami B décide de faire un cadeau à A et achète le produit P
Etape 4: A remplace le produit de la wish list par un nouveau produit, et le système reprend son processus initial.

Le modèle simple représenté par les réseaux de Petri est le suivant :

Figure 2.1



Le modèle se définit donc par les places et transitions suivantes :

Places	**Description**
P1	A est inactif. Le système est à son état intial.
P2	A a déposé sa wish list
P3	B a consulté la wish list de A
P4	B a acheté le cadeau à A sur la wish list
Transitions	
T1	A dépose sa wish list
T2	B consulte la wish list de A
T3	B achète le cadeau à A
T4	A change le produit de la wish list

Tableau 2.1

Chacune des transitions possède un état de départ et un état d'arrivée. Pour qu'une transition s'effectue, l'état de départ doit comporter au moins un jeton. Dans ce cas, lorsqu'une transition survient, suite à un événement, le changement d'état est appelé *franchissement* ou *tir* de la transition. Après cette transition, le jeton disparaît de l'état de départ et apparaît dans l'état d'arrivée . On voit sur cet exemple simple que le premier intérêt des réseaux de Petri est

d'obtenir une représentation graphique claire des étapes du phénomène. Nous verrons ultérieurement que l'analyse des propriétés du réseau nous permettra d'analyser les points de dysfonctionnement du système afin de l'optimiser (dans la figure 2.1, il n'y a pas de dysfonctionnement puisque le système marche « efficacement », le processus aboutissant à un achat final par B).

L'état global d'un réseau de Petri se définit par un ensemble de places : certaines sont « non marquées » (pas de jeton), d'autres sont « marquées » (avec un ou plusieurs jetons). La distribution correspondante des jetons dans les places est appelée *marquage* du réseau de Petri.

La représentation de la dynamique du système suit la règle suivante: la transition d'un marquage vers le marquage suivant est *franchissable* et s'effectue lorsque toutes les conditions relatives à cette transition sont satisfaites : plus précisément, le nombre de jetons de chaque place entrée d'une transition doit être supérieur ou égal au « poids » de l'arc joignant cette place à la transition (le poids affecté à l'arc est le nombre de jetons nécessaires pour que la transition suivant l'arc s'effectue): par exemple, supposons un système de wish list qui ne fonctionne que si deux amis au moins achètent un cadeau à A. Dans ce cas, deux agents B sont nécessaires pour effectuer la transition T_3 de la figure 2.1, donc 2 jetons seront nécessaires dans la place P_3 précédant la transition T_3, et l'arc entre la place P_3 et la transition T_3 aura un poids de 2. Lorsque aucune précision n'est fournie sur le graphique, le poids des arcs est égal à un par défaut. Ceci définit le *franchissement* d'une transition, qui dépend donc des jetons. Par principe, le *marquage initial* du système étudié sera défini par la distribution des jetons à l'état initial, considéré comme le début du comportement modélisé .

Le fonctionnement de base d'un réseau de Petri est illustré dans la figure 2.2 ci-dessous :

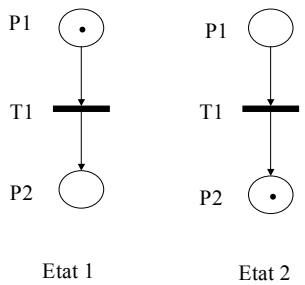


Figure 2.2

Dans la figure 2.2, la place P1 comprend un jeton et la transition T1 peut donc être franchie. Dans l'état 2, la transition T1 a été franchie, le jeton disparaît de la place P1 et apparaît dans la place P2. Ce système peut par exemple représenter l'achat d'un consommateur : le jeton représente le consommateur, l'état 1 celui-ci avant l'achat. La transition T1 est l'acte d'achat. L'état 2 correspond au consommateur ayant acheté.
Un deuxième exemple de base est illustré dans la figure 2.3 :

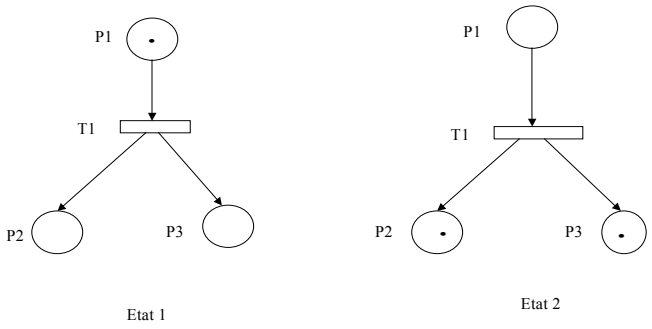


Figure 2.3

Cet exemple illustre le phénomène de dédoublement des jetons. Le franchissement de la transition T1 entraîne la création d'un jeton dans toutes les places de sorties (états d'arrivée) de la transition T1.
Après cette description des principes de base d'un réseau de Petri, reprenons maintenant le cas de la wish list décrit précédemment.
Dans la figure 1, le marquage initial du système est :

$$M_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{matrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{matrix}$$

Dynamique du système

Dans ce réseau de Petri, le franchissement de la transition T_1 (A dépose sa wish list) conduit à un nouveau marquage M_1 (nouvel état du système) qui correspond à l'état où A a déposé sa wish list. Ce nouveau marquage est obtenu à partir du marquage initial M_0 , en supprimant

d'abord, dans les places d'entrée de T_1 (donc P_1) , le nombre de jetons indiqués sur les arcs entrant de T_1 (noté Pré (P_1, T_1), dans notre exemple, il n'y a qu'un jeton passant de P_1 à T_1) et en ajoutant ensuite, à la place de sortie P_2 de T_1, le nombre de jetons correspondant au poids indiqué sur l'arc sortant de T_1 vers P_2 (noté Post (P_1,T_1), dans notre exemple, il n'y a qu'un jeton passant de T_1 à P_2). Sur la figure 1, ceci correspond (sur cet exemple simple) au fait d'enlever un jeton dans P_1 et d'en ajouter un dans P_2. La modélisation de la wish list de la figure 2.1 par un réseau de Petri permet de représenter la dynamique du système de la manière suivante:

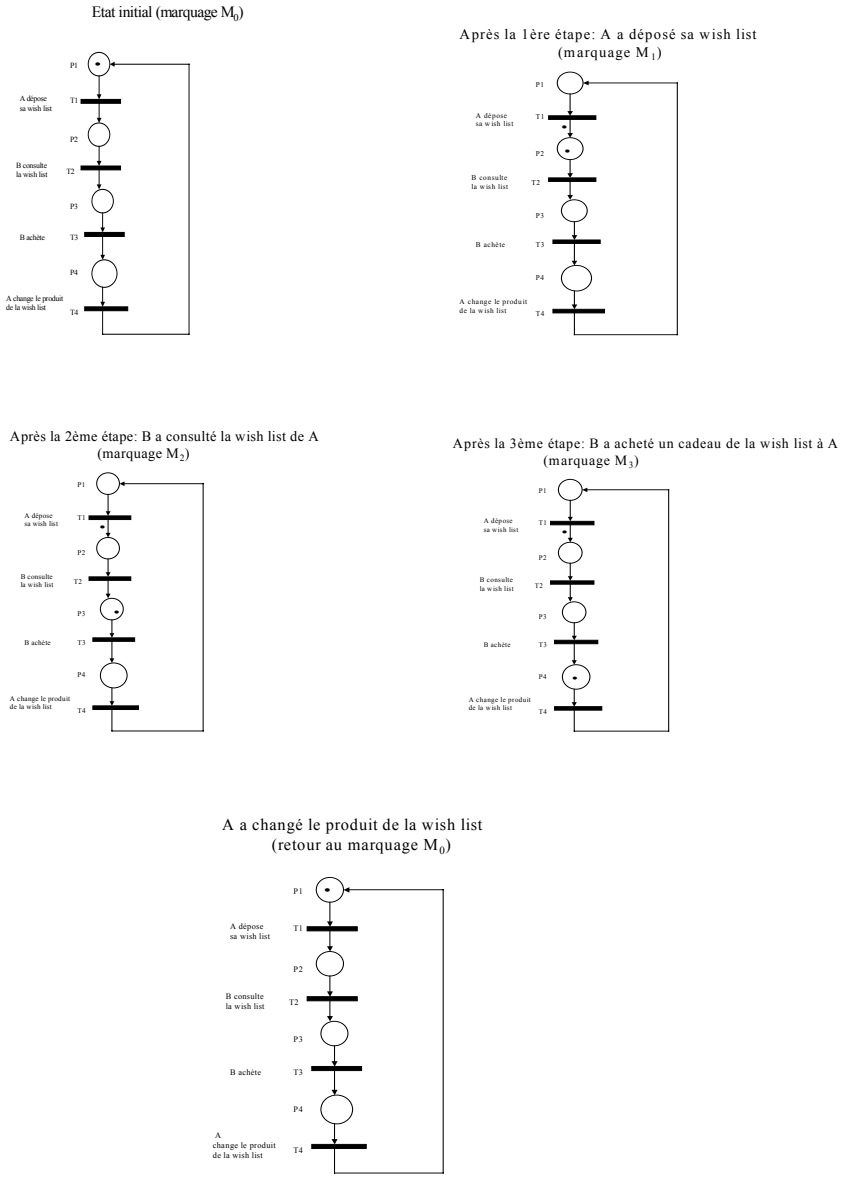


Figure 2.4 : dynamique du système

Les différents états de ce système correspondent aux marquages suivants :

$$M_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad M_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad M_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \qquad M_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \qquad M_4 = M_0$$

Les différents marquages de ce réseau de Petri sont définis en appliquant l'équation fondamentale des réseaux de Petri, à savoir :

$M_1 = M_0 + W.T_1$

$M_2 = M_1 + W.T_2$

$M_3 = M_2 + W.T_3$

Avec :

W : matrice d'incidence définie par :

$W = W^+ - W^-$ avec:

$W^+ = [w_{ij}^+]$ où $w_{ij}^+ = \text{Post}(P_i, T_j)$

$W^- = [w_{ij}^-]$ où $w_{ij}^- = \text{Pré}(P_i, T_j)$

Pré (P_i , T_j) = n (poids associé à l'arc allant de P_i à T_j, ou encore ce qui est nécessaire pour effectuer la transition T_j)

Post (P_i , T_j) = n' (poids associé à l'arc allant de T_j à la place suivante, ou encore ce qui est produit par la transition T_j)

Pour l'exemple de la wish-list décrite dans la figure 2, on a:

$$W^- = \begin{array}{c} \begin{matrix} T1 & T2 & T3 & T4 \end{matrix} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \begin{matrix} P1 \\ P2 \\ P3 \\ P4 \end{matrix}$$

$$W^{+} = \begin{array}{c} \begin{matrix} T1 & T2 & T3 & T4 \end{matrix} \\ \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array} \begin{matrix} \\ P1 \\ P2 \\ P3 \\ P4 \end{matrix}$$

$$W = \begin{array}{c} \begin{matrix} T1 & T2 & T3 & T4 \end{matrix} \\ \begin{pmatrix} -1 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix} \end{array} \begin{matrix} \\ P1 \\ P2 \\ P3 \\ P4 \end{matrix}$$

On peut donc ainsi modéliser toutes les étapes du système.

Définition mathématique d'un réseau de Petri :

Un réseau de Petri places-transitions R se définit par un quadruplet (P, T, Pré, Post), avec :

1/ $P=\{P_1,\ldots,P_i,\ldots,P_n\}$ est un ensemble fini et non vide de places

2/ $T=\{T_1,\ldots,T_j,\ldots,T_m\}$ est un ensemble fini et non vide de transitions

3/ Les ensembles P et T sont disjoints

4/ Pré : $P \times T \rightarrow N$ est l'application d'incidence avant, définie par

Pré (P_i , T_j) est la valeur entière n associée à l'arc allant de P_i à T_j

5/ Post : $P \times T \rightarrow N$ est l'application d'incidence arrière, définie par :

Post (P_i , T_j) est la valeur entière associée à l'arc allant de T_j à P_i

Equation générale d'un réseau de Petri :

Plus généralement, l'équation fondamentale d'un réseau de Petri est la suivante :

Soit S une séquence de franchissements réalisable à partir d'un marquage M_i , on définit le vecteur caractéristique de la séquence S, noté $\underline{S}$, le vecteur de dimension m (m étant le nombre de transitions du système) dont la composante j correspond au nombre de franchissements de la transition Tj dans la séquence S. Notons M_k le marquage issu de la séquence de franchissements S. L'équation fondamentale du système modélisé par les réseaux de Petri est :

$M_k = M_i + W.\underline{S}$ où :

W est la matrice d'incidence définie par :

$W = W^+ - W^-$ avec:

$W^+ = [w_{ij}^+]$ où $w_{ij}^+ = Post (P_i, T_j)$

$W^- = [w_{ij}^-]$ où $w_{ij}^- = Pré (P_i, T_j)$

Ceci signifie que tout nouveau marquage M_k sera obtenu, à partir du marquage précédent M_{k-1} en supprimant d'abord, dans la ou les places d'entrée de T (T étant la transition entre le marquage M_{k-1} et M_k) , le nombre de jetons indiqué sur les arcs entrants de T, et en ajoutant ensuite, à chaque place de sortie P de T, le nombre de jetons correspondant au poids indiqué sur l'arc sortant de T vers P).

Grâce à la matrice d'incidence et à l'équation fondamentale du modèle, les réseaux de Petri nous permettent donc de modéliser les différents changements d'état en fonction des évènements dynamiques qui interviennent dans le système.

Analyse comportementale des réseaux de Petri

Il existe deux grands types d'analyse des réseaux de Petri. La première famille comprend l'analyse de la structure du réseau et de son comportement afin de détecter des points de dysfonctionnements. Ce premier type d'analyse, relativement technique, est très utilisé en informatique, en recherche opérationnelle ou en gestion de production lors de la conception d'un système (les questions posées sont du type : où se trouvent les dysfonctionnements ? y-a-t-il des points de blocage ? le système vérifie-t-il les propriétés attendues ?).

Nous n'utiliserons pas cette première famille d'analyse au sein de la thèse, car notre problématique ne porte pas sur la conception de système mais plutôt sur la modélisation de processus , l'évaluation de ses performances et sur la simulation.
En revanche, afin d'avoir une présentation complète des réseaux de Petri, nous présentons en annexe 1 les principaux éléments de cette première famille d'analyse sur le comportement des réseaux de Petri.
La deuxième famille d'analyse consiste à étudier les performances du système et à mener des simulations. Ces techniques seront développées et appliquées dans les chapitre 3 et 4 sur les cas que nous avons modélisés dans la thèse.

2.2.3 Deuxième exemple : description du processus de commande d'un client :

On peut représenter par un réseau de Petri un processus classique en marketing : la commande d'un client, comme l'illustre la figure 2.5 ci-dessous.

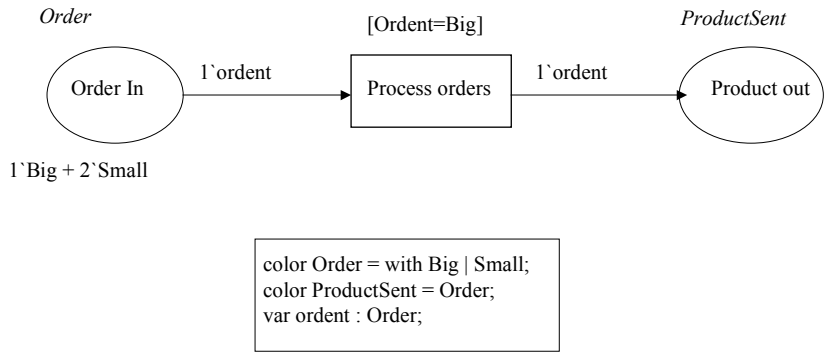


Figure 2.5

Le traitement des commandes clients d'une entreprise peut être schématisé très simplement par le réseau de Petri de base suivant : l'état initial du système (place *OrderIn)* comprend l'ensemble des commandes à traiter, la transition *ProcessOrders* correspond au traitement de la commande et la place *Product out* contient l'ensemble des commandes traitées envoyées aux clients.

Nous allons dans un premier temps expliquer les éléments de ce modèle, puis nous discuterons de son intérêt pour répondre à des questions managériales.

Ce réseau simple est modélisé à partir du logiciel Design/CPN. Les réseaux de Petri peuvent être programmés sur plusieurs types de logiciels. Nous avons choisi dans le cadre de cette présentation de travailler sur le logiciel Design/CPN, qui a été développé par Meta Software Corporation, Cambridge MA, USA, en collaboration avec des chercheurs de l'université d'Aarhus au Danemark. Actuellement, la distribution, le développement et le support de Design/CPN est effectué par le groupe CPN de l'Université d'Aarhus au Danemark. En 1996, plus de 200 institutions (universités et entreprises privées) de 30 pays différents avaient déjà utilisé cet outil (http://www.daimi.au.dk/PetriNets/). Design/CPN est l'un des logiciels les plus élaborés parmi les outils existant sur les réseaux de Petri. Ces raisons expliquent notre choix de ce logiciel (utilisable sous Linux). Actuellement, une version sous Windows est en cours de développement et sera accessible prochainement (cette version se nomme CPN Tools).

Reprenons le réseau de la figure 2.5, et expliquons le modèle, sans rentrer de manière détaillée dans la programmation (pour plus d'informations sur la programmation, on peut consulter le manuel d'utilisateur de Design/CPN téléchargeable sur le site http://www.daimi.au.dk/PetriNets/).

Les places :

Le réseau de Petri de la figure 2.5 contient deux places : *Order in* et *Product out* .

Chaque place est définie par un nom (situé au milieu du cercle) et par le type de jetons qu'elle contient. Les jetons peuvent en effet être des chiffres, du texte, des vecteurs, des valeurs booléennes, etc…. Par exemple, si les jetons identifient des types de clients différents, chaque client sera alors identifié par un vecteur contenant ses caractéristiques (socio-démographiques, historique d'achat, etc.). Dans un réseau de Petri, le type de jeton d'une place est déclaré dans la « déclaration globale », programme généralement encadré et situé en dessous du graphique du réseau (cf figure 2.5). Ce programme contient les définitions de tous les paramètres du système modélisé (langage SML).

Dans la figure 2.5, la place de gauche s'appelle *Order In* et représente l'ensemble des commandes clients. Les jetons correspondent donc aux commandes clients qui dans l'exemple sont de 2 types (Small (petites commandes) ou Big (grosses commandes)), comme cela est défini dans la déclaration globale (ligne : color order = with Big | Small).

La place de droite s'appelle *Product Out* et représente l'état du système dans lequel la commande est passée et le produit envoyé au client. Cette place contient les mêmes jetons que la place de gauche (ligne : Color ProductSent = Order ;), à savoir des commandes de deux types possible (grosses ou petites).

Marquage des places :

Comme déjà exposé dans la partie 2.2.2, une place peut contenir des jetons. Dans la figure 2.5, l'expression 1`Big + 2`Small, associée à la place *Order In* signifie que cette place contient comme marquage initial un jeton de valeur Big et deux jetons de valeur Small. Par conséquent, au départ du système, cette place contient une grosse commande et deux petites commandes.

La place *Product Out* n'a pas de marquage initial, ce qui signifie qu'à l'état initial du système, cette place ne contient aucun jeton.

Nous détaillerons ultérieurement la dynamique du système de la figure 2.5 et l'évolution de son marquage, afin de montrer l'intérêt d'utiliser les réseaux de Petri dans la gestion des commandes clients.

Les variables du réseau de Petri :

La simulation d'un réseau de Petri requiert souvent de définir les jetons utilisés comme des variables. Ces variables doivent être déclarées, chacune des variables étant d'un type particulier et pouvant prendre uniquement les valeurs correspondant à ce type. Le réseau de Petri de la figure 2.5 comporte un seul type de variable: *ordent,* qui peut prendre soit la valeur Big soit la valeur Small (ceci correspond à la ligne : var ordent : Order). Ce sont les seules valeurs que cette variable peut avoir.

Les transitions :

Une transition dans un réseau de Petri est un événement qui arrive dans le système et qui change le nombre ou la valeur des jetons dans les places. Elle est représentée généralement par un rectangle à l'intérieur duquel est inscrit son nom. Le réseau de Petri de la figure 2.5 ne contient qu'une transition, *Process Orders*, qui représente le processus de prise en charge de la commande client. A cette transition est associée une condition à son franchissement. Cette condition est exprimée par l'expression: [ordent=Big], signifiant que cette transition ne peut s'effectuer que si la place d'input de cette transition (*Order In*) contient un jeton égal à Big. Dans le réseau de la figure 5, seules les grosses commandes sont traitées.

Les inscriptions sur les arcs :

Le réseau de Petri de la figure 2.5 a deux arcs, l'un reliant la place Order In à la transition Process Orders, l'autre reliant la transition Process Orders à la place Product Out.

Une inscription sur un arc est une expression associée à cet arc. Sur le réseau de Petri de la figure 2.5, les deux arcs portent la même inscription : *1`ordent*. Ceci signifie que les jetons spécifiés sur l'inscription de l'arc (donc des jetons de type Big ou Small) doivent être présents dans la place d'entrée de la transition pour que celle-ci soit franchissable. Lorsque la transition s'effectue, ces jetons sont supprimés de la place d'entrée. L'inscription de l'arc de sortie de la transition définit les jetons qui seront ajoutés à la place de sortie (dans la figure 2.5, seuls les jetons de type Big ou Small transiteront vers la place *Product Out*).

Dynamique d'un réseau de Petri :

Nous avons jusqu'ici décrit les différents éléments du réseau. Une fois programmé sur le logiciel Design CPN, le réseau de Petri apparaît sous la forme graphique suivante, permettant ainsi la phase de simulation :

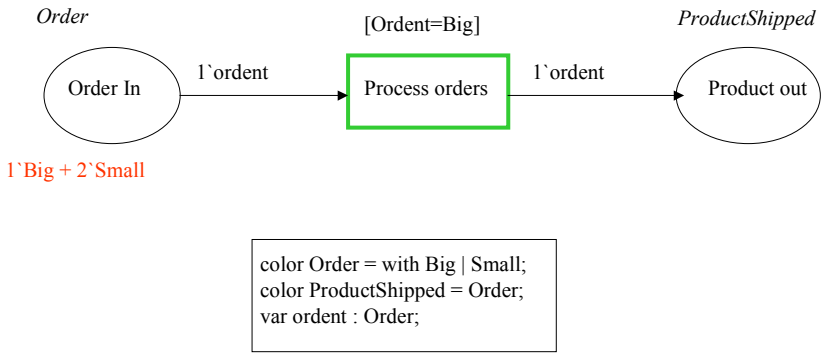


Figure 2.6

Lorsqu'on rentre dans le menu simulation de design CPN, la transition Process Orders apparaît en vert, signifiant que la transition est franchissable (en effet, nous avons dans la place d'entrée Order In un jeton qui vérifie bien la condition de la transition (ordent = Big)).

On simule alors le système : la transition Process Orders s'effectue et le jeton Big est supprimé de la place Order In et ajouté à la place de sortie Product Out. Ceci signifie que la grosse commande du client à été traitée et envoyée à ce client.
L'état du système devient donc :

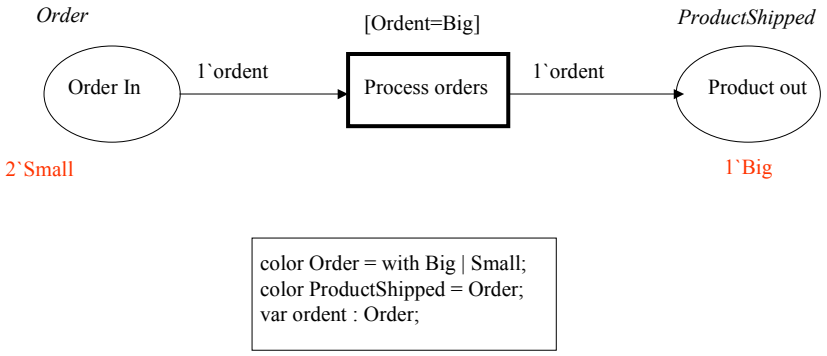


Figure 2.7

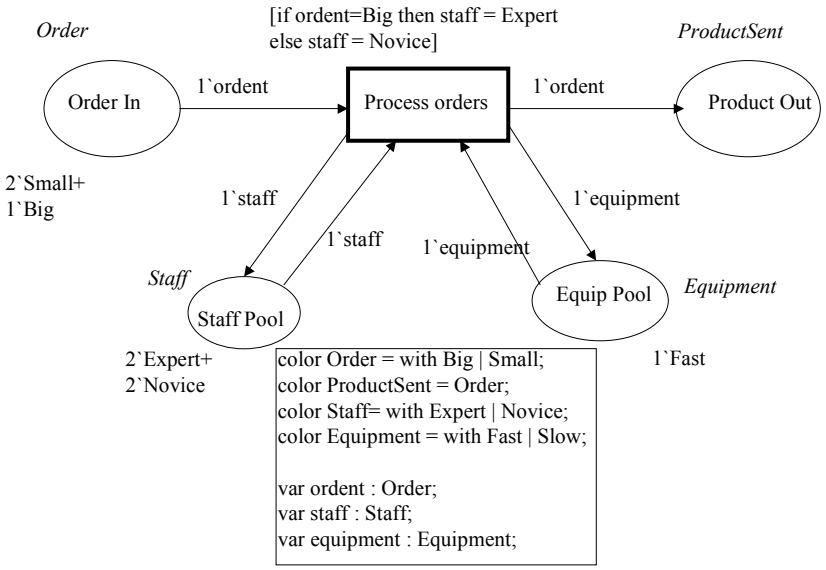
Le nouvel état du système est alors le suivant: une grosse commande a été traitée et envoyée (on a un jeton Big dans la place Product Out), et deux petites commandes n'ont toujours pas été traitées (il reste 2 jetons Small dans la place Order In). Le système s'arrête ensuite puisqu'il n'y a que des petites commandes dans la place d'entrée Order In et qu'elles ne respectent pas la condition de la transition (On peut par exemple penser à un système dans lequel on traite immédiatement les grosses commandes et 2 jours après les petites commandes). La transition Process Orders ne peut pas être franchie, c'est la raison pour laquelle elle apparaît en noir. Le système est dans son état final.

Extensions du réseau et analyse d'un système plus complexe :

Nous avons jusqu'ici montré les points essentiels d'un réseau de Petri et de Design/CPN. En revanche, nous n'avons pas donné d'éclairage sur la manière dont les réseaux de Petri peuvent être utilisés pour modéliser des systèmes réels.

Nous allons par conséquent dans cette partie étendre et complexifier légèrement le réseau de Petri simple de la figure 2.5 afin de traiter un cas plus proche de la réalité qui permettra de comprendre les avantages des modèles Petri.

Figure 2.8



Ce système décrit le traitement de commandes clients mais détaille davantage les ressources utilisées par l'entreprise pour traiter ces commandes. Les places d'entrée et de sortie sont les mêmes que celles de la figure 2.5 et ont été expliquées précédemment. Lorsqu'une commande est traitée, elle utilise des ressources humaines (définies dans la place *StaffPool*) et des équipements (place *Equip Pool*). Selon le type de commande (*Big* ou *Small*), l'entreprise utilise un personnel spécifique : soit des experts, soit des novices. A l'état initial du système, le personnel disponible dans la place *StaffPool* est constitué de 2 experts et de 2 novices. Le traitement des commandes demande également l'utilisation d'un équipement particulier. L'entreprise en possède deux types : les équipements rapides (*Fast*) ou lents (*Slow*). A l'état initial du système, l'entreprise a un seul équipement disponible (un rapide (*fast*) défini dans la place *EquipPool)*.. Les grosses commandes sont traitées différemment des petites commandes : les grosses commandes sont traitées par un expert tandis que les petites commandes sont traitées par un novice (cette condition est associée à la transition *Process orders :* ceci correspond dans la figure 8 à la notation: [if ordent=Big then staff = Expert else staff = Novice]). Généralement,, dans les réseaux de Petri, une condition associée à une transition est appelée une *garde*.

En revanche, il n'y a pas de condition pour l'équipement : une grosse ou petite commande peut être traitée par un équipement rapide ou lent.

Le modèle représentant ce système est paramétré avec tous les éléments nécessaires dans la figure 8.

Parmi les questions manageriales qui peuvent se poser, citons par exemple les suivantes : que se passe-t-il lorsque de nombreuses commandes arrivent, lorsque plusieurs personnes sont en charge de ces commandes, lorsque plusieurs équipements sont utilisés ? Existe-t-il des moments de blocage du système, où une commande n'est pas traitée? Quelle serait la meilleure solution d'organisation pour traiter tant de grosses commandes, tant de petites commandes ?

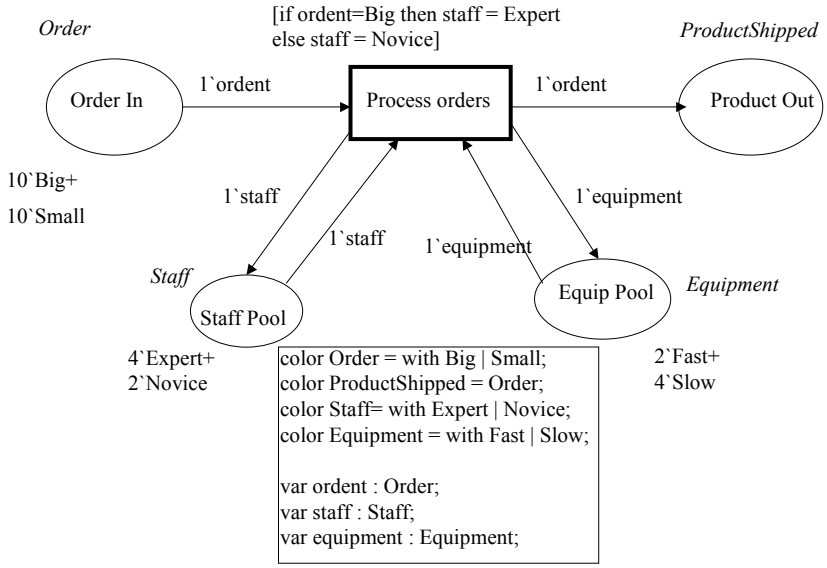
Lorsque le système devient complexe, les réponses à ces questions ne sont pas triviales. Les réseaux de Petri permettent de simplifier cette complexité et de pouvoir répondre à ces problématiques complexes. Pour cela, on regarde le fonctionnement dynamique du réseau grâce au logiciel Design CPN, on analyse son graphe d'état (cf annexe 1) et on effectue des simulations afin de trouver les situations d'organisation les meilleures, ou de prévoir les flux de traitements des commandes.

A titre d'exemple, étudions la question suivante : comment le système répondrait-il en cas de fortes ventes, s'il y avait plus de commandes que de personnel pour les traiter, et une mauvaise adéquation entre le personnel et l'équipement ?

Considérons le système suivant : la place *Order In* comporte 10 grosses commandes et 10 petites commandes, la place *StaffPool* contient 4 experts et 2 novices, et la place *Equip Pool* contient 2 équipements rapides et 4 lents.

La représentation graphique de ce système à son état initial est le suivant :

Figure 2.9



Lorsque ce système s'exécute, il n'y a pas suffisamment de personnel ou d'équipements pour traiter tous les ordres. En raison de la surcharge, certaines commandes sont donc obligées d'attendre leur tour, ou ne seront peut être jamais traitées. En fait, les petites commandes resteront plus longtemps dans la place *Order In* que les grosses commandes puisqu'il y a plus de personnel *Expert* que de personnel *Novice*. On peut alors simuler différentes composition de personnel et d'équipements, en comparer les résultats pour en tirer des recommandations managériales en terme d'organisation.

Activités concurrentes :

L'ensemble des réseaux que nous avons présentés précédemment s'exécutent de manière séquentielle : un seul événement arrive à un moment donné. Une telle exécution ne nécessite aucune sophistication particulière ou complexité de simulation et d'analyse. Il exécute une action, puis une autre, puis la suivante, en fonction des règles programmées dans le système (s'il n'y a pas de règle, le système choisit dans les places d'entrée des transitions franchissables un jeton de manière aléatoire), jusqu'à ce qu'il n'y ait plus aucune action possible au sein du système. Cependant, les opérations séquentielles ne correspondent pas à la majorité des systèmes dans la réalité. La modélisation d'un système doit intégrer toutes les actions qui peuvent être effectuées au même moment. Les évènements qui arrivent au même

moment sont appelées des activités concurrentes. Les réseaux de Petri sont particulièrement performants pour modéliser des systèmes concurrents.
Reprenons et complexifions le réseau décrit dans la figure 2.9.
Considérons la figure 10 qui modélise une entreprise qui traite les commandes selon le cycle d'actions suivantes pour chaque commande client :

1. Une commande rentre dans le système pour être traitée (transition *EnterOrder*).
2. Une personne (issue d'une équipe représentée par la place *StaffPool)* est assignée à prendre en charge cette commande. Si la commande est grosse, cette personne est un expert, si elle est petite, la personne est novice (ceci est indiqué dans la garde de la transition *EnterOrder*).
3. La personne obtient une pièce d'équipement (parmi l'ensemble des pièces d'équipement de la place *EquipPool*), l'utilise pour traiter la commande et retourne l'équipement (qui revient dans la place *EquipPool*). Une grosse commande exige un équipement rapide, une petite commande requiert un équipement lent (cf garde de la transition *ProcessOrder*).
4. La personne expédie le produit (Transition *SendProduct*).
5. La personne est immédiatement disponible pour traiter une autre commande (elle retourne dans la place *StaffPool*).
6. Le traitement de la commande est complet.

Exemple :

Supposons que le marquage initial du ce réseau soit le suivant :

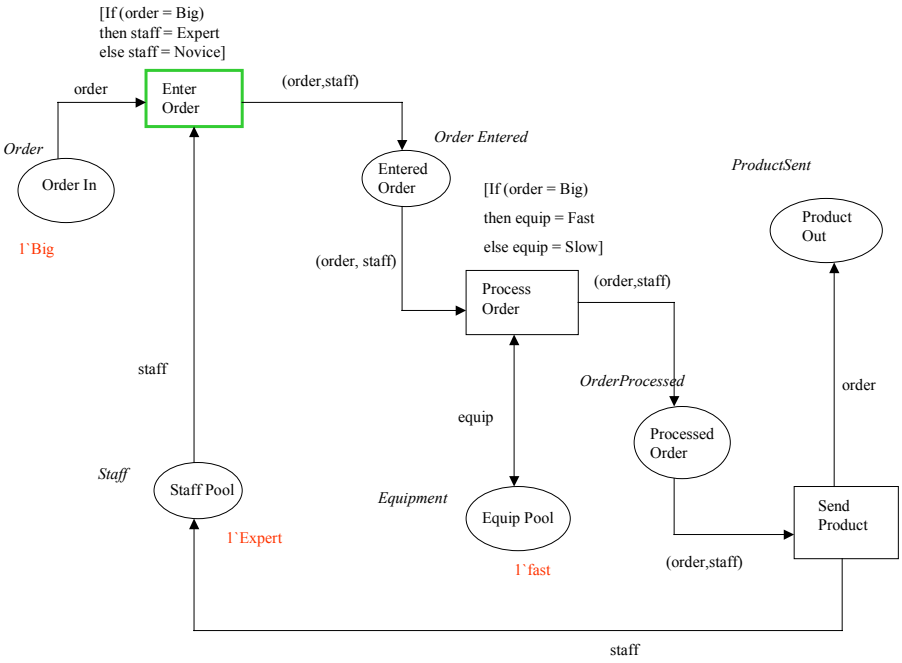


Figure 2.10

La réception de la commande sera modélisée par :

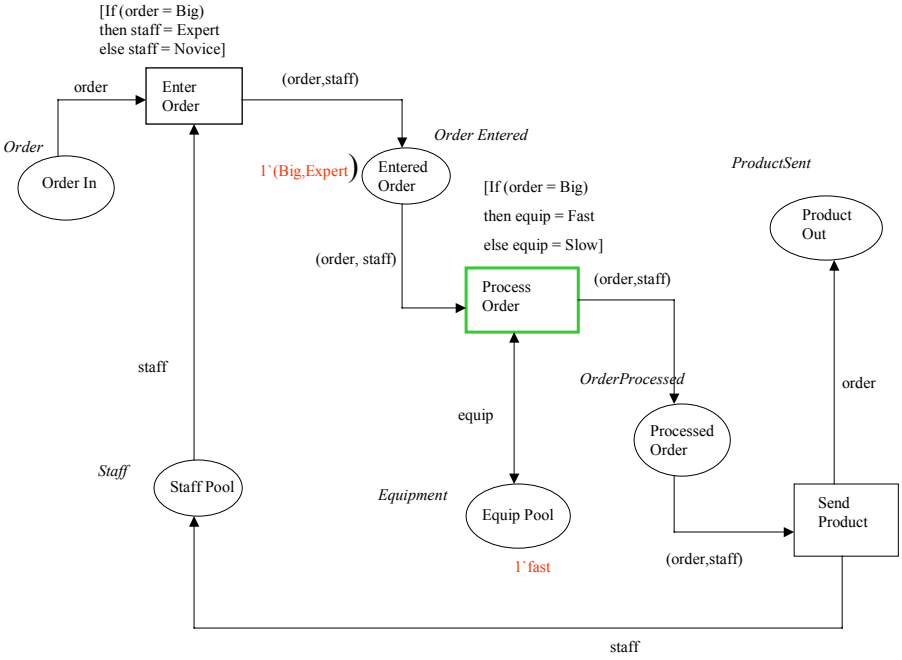


Figure 2.11

Le traitement de la commande par :

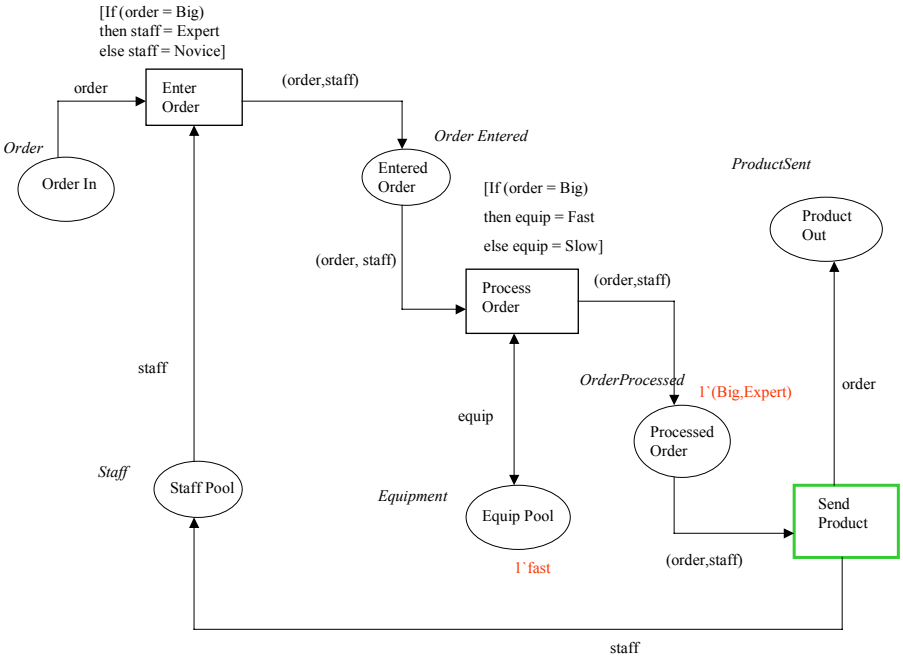


Figure 2.12

L'expédition de la commande par :

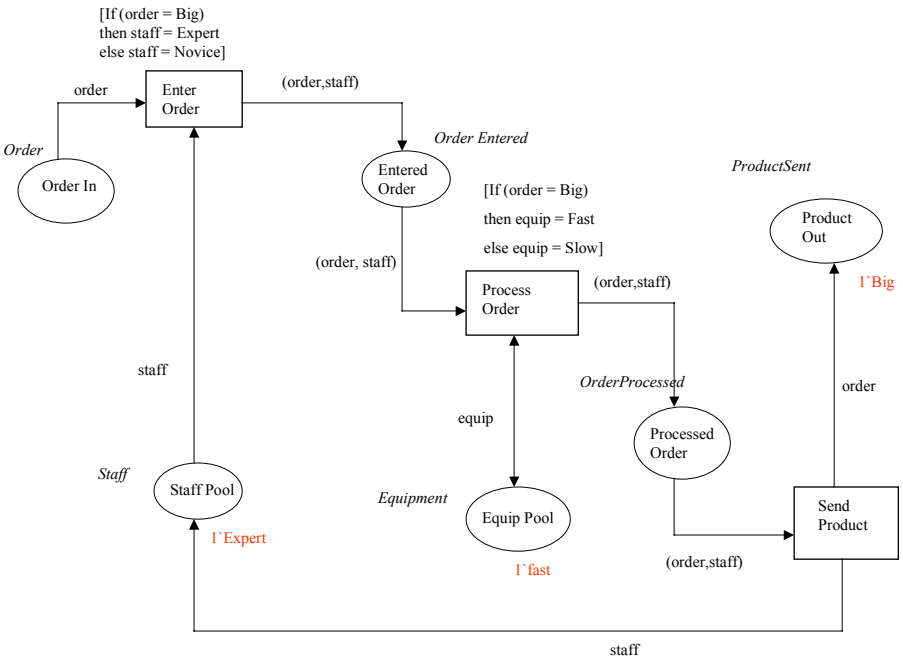


Figure 2.13

Nous avons présenté un exemple simple avec une seule commande, un seule personne (*Expert*) en charge, et une seule pièce d'équipement rapide (*Fast*). L'intérêt des réseaux de Petri est d'analyser le système avec plusieurs commandes, plusieurs experts et novices, plusieurs équipements. Dans ce cas, plusieurs tâches concurrentes sont effectuées en même temps (une commande est enregistrée tandis qu'une autre commande est en cours de traitement, en même temps que plusieurs commandes sont en train d'être expédiées). Les réseaux de Petri sont particulièrement puissants pour modéliser ce type de situation car ils fournissent la possibilité de modéliser plusieurs processus s'exécutant en même temps, avec plusieurs agents différents intervenant dans le système. Dans le réseau de la figure 2.13, lorsqu'il y a plusieurs commandes, plusieurs personnes en charges et plusieurs équipements, il est impossible de prévoir sans modélisation ce qui va se passer dans le système. Les réseaux de Petri nous permettent de modéliser cette complexité.

Lorsque des douzaines de transitions s'exécutent simultanément, on peut dans Design/CPN contrôler le système. On peut par exemple voir comment s'exécute chaque étape dans la simulation. On peut ainsi regarder à la loupe le fonctionnement du système, le « debugger » s'il y a des points de blocage (ceci s'applique lorsqu'on construit un système) ou simuler ce qui se passe avec un nombre important de données (dans notre exemple, que se passe-t-il avec 1000 commandes, 100 experts, 20 novices et 48 équipements ?). On peut ainsi savoir

exactement ce qui se passe dans le système, le modifier pour l'améliorer et faire en sorte qu'il corresponde au système recherché.

2.2.4 Les différentes familles de réseaux de Petri :

Les nombreuses recherches menées sur les réseaux de Petri, les cas d'applications et le courant de recherche très actif sur ce type de modélisation ont entraîné le développement de plusieurs familles issues du modèle de base, comme le montre la figure 2.14 ci-dessous :

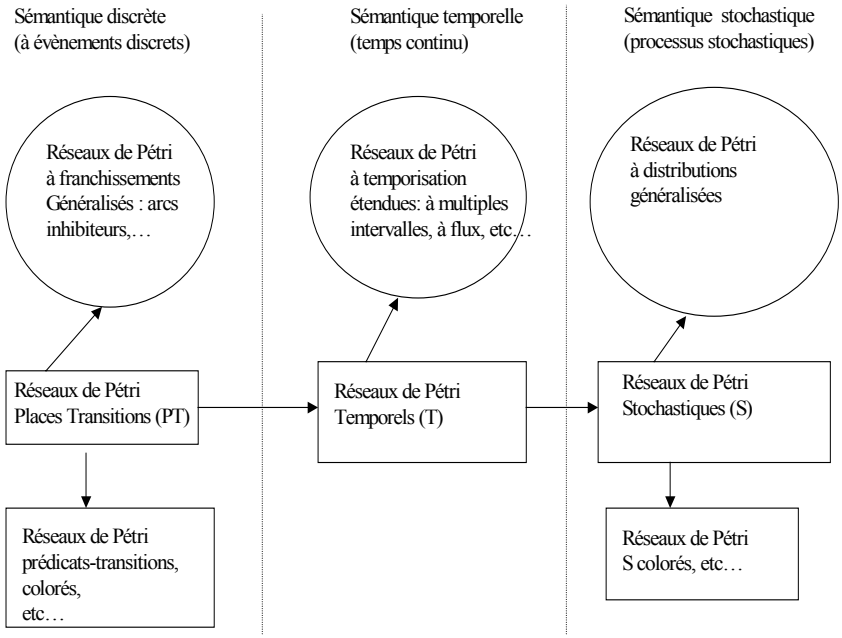


Figure 2. 14: Les domaines sémantiques des réseaux de Pétri
Source: Diaz (« Les réseaux de Pétri, modèles fondamentaux », 2001)

On constate sur la figure 2.14 l'existence de trois grands domaines:

1. La famille des réseaux de Petri à sémantique discrète, dont le modèle de base est le réseau de Petri places-transitions (une succession finie de places et transitions immédiatement franchissables, avec possibilité de fixer des conditions aux transitions). Les comportements de ces systèmes peuvent se représenter par des graphes (cf annexe 1) finis ou dénombrables d'états. Le réseau de Petri de la figure 10 modélisant le traitement des commandes clients est un réseau de Petri places-transitions.

Les extensions des réseaux de Petri places-transitions sont les réseaux de Petri à franchissement généralisés (comprenant par exemple les arcs inhibiteurs, qui seront définis dans la partie 2.3 et qui permettent le test à zéro d'une place) ou les réseaux colorés, dans lesquels les jetons peuvent être différenciés par des couleurs (paramétrés par des vecteurs), ce qui permet de modéliser des systèmes complexes (par exemple en marketing pour différencier plusieurs types de clients, ou plusieurs produits).

2. La famille des réseaux de Petri à sémantique temporelle, qui tient compte du temps, dont le modèle de base est le réseau de Petri Temporel qui associe un intervalle de temps au franchissement des transitions (la transition ne peut être franchie que dans cet intervalle de temps). Des modèles issus du modèle temporel de base permettent de représenter des systèmes plus complexes incorporant la notion de temps.
3. La famille des réseaux de Petri stochastiques : cette famille fait l'objet de la partie 1.3 et consiste à modéliser des transitions dont le délai de franchissement suit un processus stochastique (processus exponentiel dans le modèle de base). Les réseaux de Petri stochastiques ont donné naissance ensuite aux réseaux de Petri stochastiques colorés, dans lesquels les jetons peuvent être différenciés, ou aux réseaux de Petri à distributions généralisées, dans lesquels les délais de franchissement des transitions peuvent suivre un nombre varié de type de distributions (Gamma, Weibull, Hyper exponentielle, Log normale, …). Nous appliquerons cette dernière famille dans les chapitre 3 et 4 de cette thèse.

Notre objectif n'étant pas d'effectuer une bibliographie complète sur les réseaux de Petri, nous nous limiterons à la présentation ci-dessus en ce qui concerne la description des différentes familles de modèles. Pour plus d'approfondissement, on peut se référer à Diaz (2001), qui consacre un chapitre à chaque famille de réseau de Petri présentée ci-dessus.

En revanche, nous allons développer dans la partie 2.3 ci-dessous la famille des réseaux de Petri stochastiques, car nous utiliserons plus particulièrement cette famille de réseaux de Petri dans les cas d'applications de la thèse.

2.3 LES RESEAUX DE PETRI STOCHASTIQUES

2.3.1 Présentation :

Introduction

Un réseau de Petri stochastique est obtenu à partir d'un réseau de Petri classique en associant une fonction de distribution de probabilité pour le délai de franchissement de chaque transition. A chaque transition sera donc spécifiée une loi de probabilité déterminant le délai selon lequel la transition va supprimer des jetons dans ses places d'entrée et les ajouter dans ses places de sortie. Un réseau de Petri stochastique sera noté SPN (Stochastic Petri Net). Dans le modèle des réseaux de Petri stochastique généralisés (GSPN), seulement deux distributions sont autorisées: exponentielle et déterministe de valeur 0. Les transitions qui ont une distribution exponentielle associée sont dites « temporisées » ou « timed » ; les transitions avec une distribution déterministe de valeur temps zéro sont dites « immédiates ». Dans les réseaux de Petri stochastiques étendus, les transitions sont classées de la même manière, mais une distribution arbitraire peut être associée à chaque transition temporisée. Les transitions peuvent avoir l'une des distributions suivantes : constante, uniforme, géométrique, exponentielle, Weibull, Gamma, normale, log-normale, Erlang ou une distribution hyper-exponentielle. Ceci complique les analyses de comportement du réseau et les méthodes de simulation utilisées. Lorsque la distribution de la loi de franchissement des transitions et soit exponentielle, soit immédiate, le réseau de Petri stochastique (SPN) est dit Markovien. Lorsque les transitions sont caractérisées par l'un des types de distribution générales énumérés précédemment, le réseau de Petri stochastique (SPN) est dit Non Markovien. Les SPN Non Markoviens constituent un outil très puissant pour modéliser des systèmes complexes et des processus non stationnaires, ce qui représente un avantage considérable par rapport aux modèles classiques utilisant des chaînes de Markov.

D'autres éléments sont propres aux réseaux de Petri stochastiques, que nous allons décrire ci-dessous.

Priorité :

Dans certains types de système, il est nécessaire de pouvoir représenter des évènements qui seront prioritaires par rapport à d'autres. Par exemple, on peut imaginer un système de gestion de clients pour lequel les demandes des gros clients sont toujours traitées prioritairement par

rapport à celles de petits clients. Pour modéliser ce type de contrainte, on utilise dans les réseaux de Petri stochastiques la notion de *priorité*. Le principe est d'associer à chaque transition une *priorité* qui permettra de sélectivement empêcher le franchissement de certaines transitions et de privilégier le franchissement d'autres transitions. Si S est l'ensemble des transitions franchissables lors d'un certain marquage du réseau, et si la transition avec la plus forte priorité est k, alors toute transition de S avec une priorité inférieure à celle de la transition k ne pourra être franchie.

Arc inhibiteur :

Un autre moyen de bloquer le franchissement d'une transition est d 'utiliser un *arc inhibiteur*. Un arc inhibiteur reliant une place *p* à une transition *t* avec une *multiplicité m* bloquera le franchissement de *t* pour tout marquage dans lequel *p* contient au moins *m* jetons.

Dans le cas classique où m=1, une transition est donc franchissable si chaque place d'entrée normale de cette transition contient au moins un jeton et si chaque place d'entrée reliée à cette transition par un arc inhibiteur possède zéro jeton. Un arc inhibiteur est représenté graphiquement par un petit cercle vide à la liaison arc-transition, et par le fait qu'il ne comporte pas de flèche, à la différence d'un arc normal.

L'exemple de la figure 2.14 permet de visualiser un exemple concret d'arc inhibiteur. Supposons un système dans lequel les gros clients sont toujours traités avant les petits clients. Dans la figure 2.14, les jetons de la place P2 représentent des gros clients et ceux de la place P1 correspondent à des petits clients. Tant que la place P2 possède un jeton (ceci signifiant qu'un gros client doit être traité) et que la transition T2 n'est pas effectuée (ce gros client n'est pas traité), le petit client de la place P1 ne peut avoir sa commande traitée. En effet, l'arc inhibiteur reliant P2 à T1 empêche le franchissement de la transition T1 tant qu'un jeton est toujours dans la place P2.

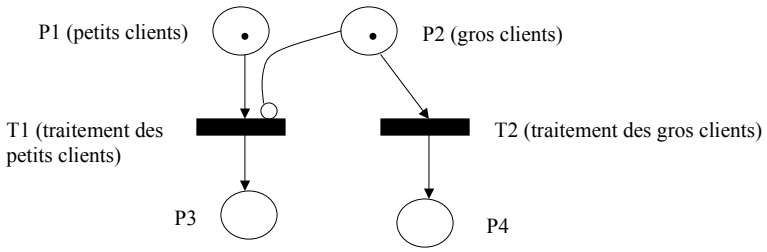


Figure 2.15: arc inhibiteur

Multiplicité des arcs et « marking dependence » :

La *multiplicité* d'un arc est un nombre associé à cet arc qui correspond au nombre de jetons que cet arc crée entre une place d'entrée et la transition suivante ou entre une transition et sa place de sortie. Par défaut la multiplicité d'un arc est égal à 1.

La figure 2.15 donne un exemple simple de multiplicité des arcs. On souhaite analyser le comportement d'achat d'un groupe de consommateurs et représenter le nombre d'items achetés à chaque acte d'achat. Dans l'état 1, le jeton de la place P1 représente un consommateur avant l'achat. On souhaite dans le système récupérer dans la place P2 le nombre d'items achetés et dans une place P3 le nombre d'actes d'achat. La transition T1 représente l'acte d'achat. La figure 25 représente deux arcs issus de la transition T1 : l'un a une multiplicité de 3 (les arcs de multiplicité supérieure à 1 sont colorés en bleu) reliant T1 à la place P2 ; l'autre de multiplicité égale à 1 reliant T1 à la place P3. Ceci signifie que lorsque le consommateur issu de P1 achète, 3 items sont enregistrés dans P2 et un acte d'achat est enregistré dans P3. La transition T1 est franchie (état 2), le jeton de la place P1 disparaît et 3 jetons sont crées dans la place P2 tandis qu'un jeton est crée dans la place P3. Le cas présenté ici ne porte que sur un consommateur dans P1, mais on peut imaginer un système plus complexe avec plusieurs centaines de consommateurs arrivant de manière dynamique (avec un délai d'arrivée suivant un processus stochastique par exemple) dans un magasin pour

acheter différents produits. Dans ce cas, les réseaux de Petri stochastiques représentent simplement ce processus plus complexe.

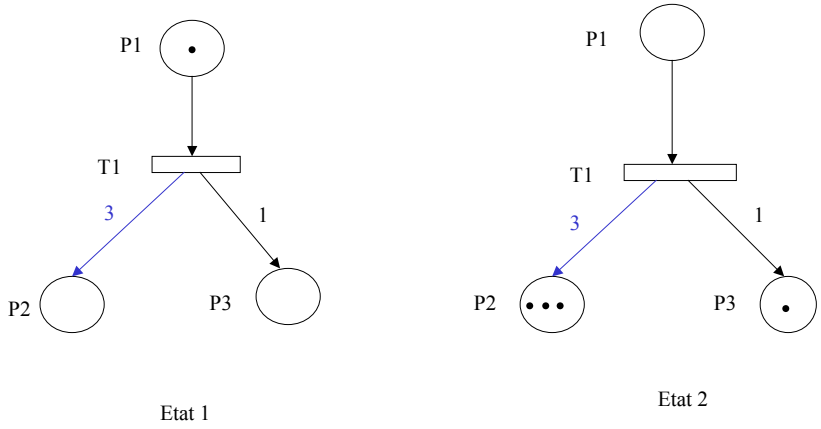


Figure 2.15: multiplicité des arcs

Dans certains cas, les arcs peuvent avoir une *multiplicité* qui n'est pas constante et qui est fonction du marquage du réseau. Cette souplesse de paramétrisation des réseaux de Petri stochastiques permet de représenter des systèmes particulièrement complexes.

2.3.2 Eléments de base d'un SPN et exemple

Les éléments de base d'un réseau de Petri stochastiques sont :

- Un ensemble fini de places $D=\{d_1, d_2, \ldots, d_L\}$
- Un ensemble fini de transitions $E=\{e_1, e_2, \ldots, e_M\}$
- Un ensemble (qui peut être vide) de transitions immédiates $E' \subset E$
- Des ensembles $I(e),J(e) \subseteq D$ de places d'entrée, de places d'entrée et de sorties, respectivement pour chaque transition $e \in E$.

Les transitions E-E' sont appelées des transitions temporisées.

G est l'ensemble des marquages du réseau.

Pour $s \in G$, on note $s= (s_1, s_2, \ldots, s_J, \ldots, s_L)$ où s_J est le nombre de jetons de la place $d_J \in D$.

Exemple illustratif : Un système producteur-consommateur :

Considérons un système constitué de deux producteurs, deux consommateurs et deux zones de stockage, numérotées 1 et 2. Les producteurs partagent une même chaîne pour transmettre les produits aux consommateurs. Le producteur 1 crée des produits pour le consommateur 1 à une date donnée. Le producteur 2 crée des produits pour le consommateur 2. Les produits crées mais non encore envoyés sont placés dans la zone de stockage du producteur i (i=1 ou 2) pour ensuite être transmis au consommateur. La zone de stockage du producteur i a une capacité finie Bi > 0. Ceci signifie qu'un produit crée par le producteur i lorsque le système contient Bi-1 produits pour le consommateur entraîne l'arrêt momentané de la fabrication de produits pour ce consommateur. Cet arrêt dure jusqu'à ce qu'un produit de la zone de stockage du producteur i soit transmis au consommateur. La paire producteur 1 consommateur 1 n'a pas de priorité sur la paire producteur 2 consommateur 2 pour l'utilisation de la chaîne de transmission. Les produits crées par le producteur i sont transmis dans l'ordre selon lequel ils sont crées. Les temps successifs requis par le producteur i pour créer un produit sont identiquement et indépendamment distribués selon une variable aléatoire Li avec une fonction de distribution continue.
Ce système peut donc être spécifié comme un réseau de Petri stochastique avec des transitions temporisées et probabilisées et des transitions immédiates.
Représentons le système par le réseau de Petri suivant :

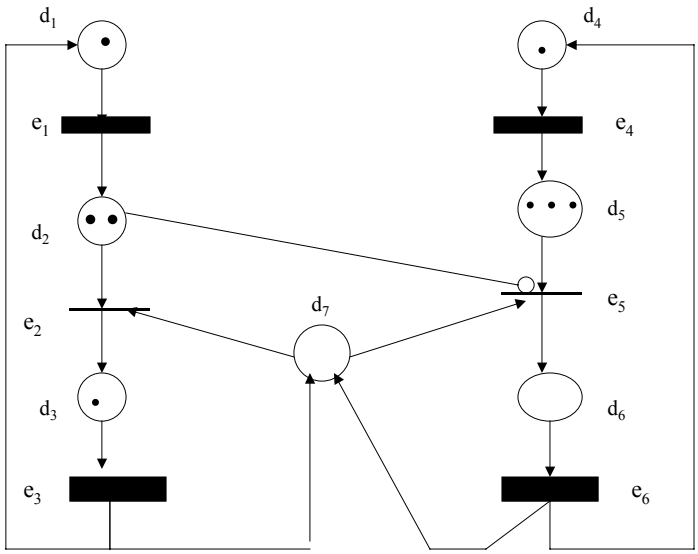


Figure 2.16: SPN représentant un système producteur-consommateur

e_1 = fabrication d'un produit par le producteur 1

e_2 = début de la transmission au consommateur 1

e_3 = fin de la transmission au consommateur 1

e_4 = fabrication d'un produit par le producteur 2

e_5 = début de la transmission au consommateur 2

e_6 = fin de la transmission au consommateur 2

Analysons la figure 2.16 avec B_1= 4 et B_2 = 3.

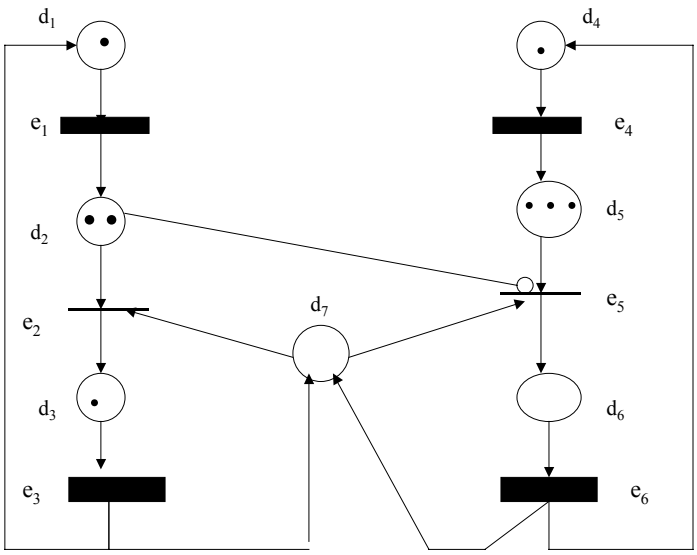
Ce réseau est spécifié par un ensemble de place D = {d_1, d_2,…., d_7}, un ensemble de transitions E = {e_1, e_2,…., e_6}et E' = {e_2, e_5} l'ensemble des transitions immédiates. La signification des transitions est décrite au dessous du graphique. L'interprétation des places est la suivante : la place d_1 contient au moins un jeton si et seulement si le producteur 1 crée un produit. La place d_3 contient au moins un jeton si et seulement si la transmission du produit au consommateur 1 est en cours. De la même manière, il y a au moins un jeton dans la place d_4 si et seulement si le producteur 2 crée un produit pour le consommateur 2 et un jeton dans la place d_6 si et seulement si la transmission d'un produit au consommateur 2 est en cours. La place d_2 (respectivement d_5) correspond à la zone de stockage du producteur 1 (respectivement 2) et contient k ($k \geq 0$) jetons si et seulement si k produits ont été fabriqués et sont en attente d'être transmis. La place d_7 modélise la chaîne de transmission commune. Elle

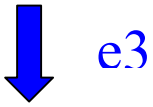
contient un jeton si et seulement si aucune transmission n'est en cours ; dans le cas contraire, la place d_7 ne contient aucun jeton.

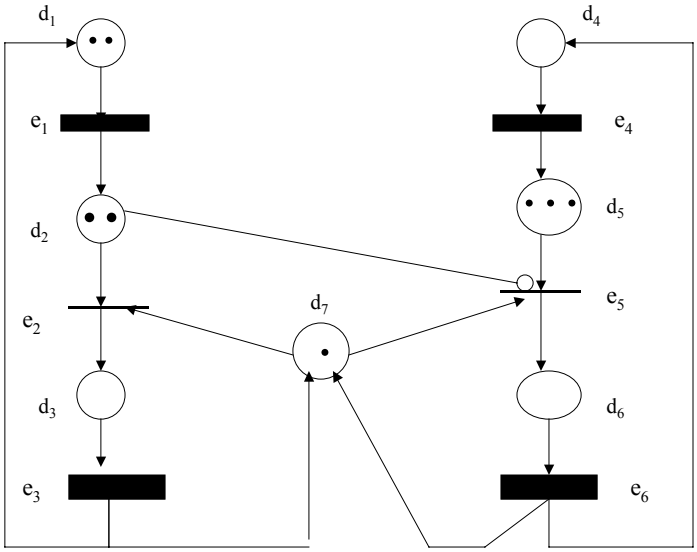
Pour mieux comprendre cette modélisation, revenons à la figure 2.16 : le producteur 1 est en train de créer un produit pour le consommateur 1, il est aussi en cours de transmission d'un produit au consommateur 1 et deux produits sont en attente de transmission pour le consommateur 1. Pour le consommateur 2, 3 produits sont en attente de transmission.

L'intérêt premier des réseaux de Petri est de pouvoir modéliser la dynamique d'un tel système. Une illustration est fournie dans la figure 2.17 ci-dessous:

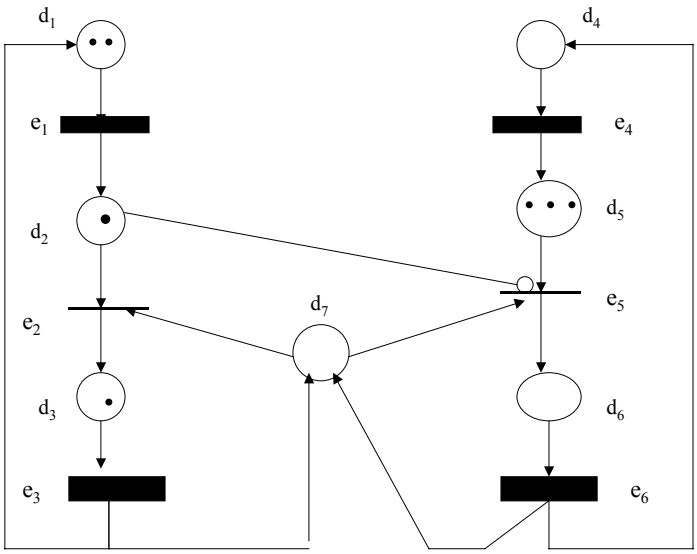
Figure 2.17: Changements de marquage représentant la dynamique du système producteur-consommateur






d1
d4
e1
e4
d2
d5
e5
d7
e2
d3
d6
e3
e6

e2


d1
d4
e1
e4
d2
d5
e5
d7
e2
d3
d6
e3
e6

La séquence des changements de marquage de la figure 2.17 montre que le modèle SPN capture bien la non priorité de la paire producteur-consommateur 1 sur la paire producteur-consommateur 2 dans l'utilisation de la chaîne globale. En effet, lorsque la transition e_3 (fin de la transmission au consommateur 1) est franchie, un jeton est déposé dans la place $d_{7,}$ indiquant que la chaîne est disponible pour la transmission d'un produit. Comme le montre la figure 2.17, des produits sont en attente de transmission pour le consommateur 1 et pour le consommateur 2 ; ces produits sont représentés par les jetons des places d_2 et d_5 respectivement. La présence de jetons dans la place d_2 entraîne le franchissement immédiat de la transition e_2 (début de la transmission au consommateur 1) empêchant simultanément le franchissement de la transition immédiate e_5 (début de la transmission au consommateur 2) grâce à l'arc inhibiteur reliant d_2 à e_5. Lorsque la transition e_2 est franchie, un jeton est déposé dans la place d_3 et une transmission au consommateur 1 débute. La transition e_2 enlève aussi un jeton de la place $d_{7,}$ indiquant que la chaîne est libre.

On peut compliquer ce système en supposant maintenant que la paire producteur-consommateur 1 a priorité sur la paire producteur-consommateur 2. Ceci signifie que lorsqu'une transmission est en cours pour le consommateur 2 et que le producteur 1 crée un produit, la transmission au consommateur 2 stoppe immédiatement et débute la transmission au consommateur 1. La figure 2.18 ci-dessous représente ce phénomène :

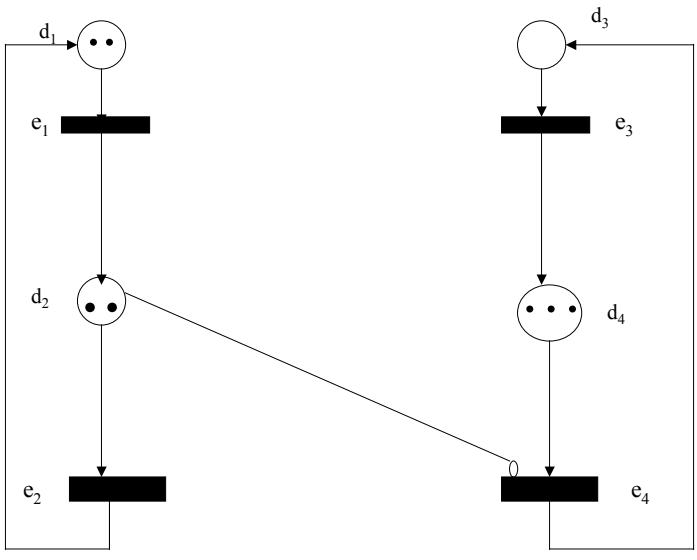


Figure 2.18: Priorité de la paire producteur-consommateur 1

e_1 = fabrication d'un produit par le producteur 1

e_2 = fin de la transmission au consommateur 1

e_3 = fabrication d'un produit par le producteur 2

e_4 = fin de la transmission au consommateur 2

Comme le montre la figure 2.18, la priorité de la paire producteur-consommateur 1 est modélisée par le fait de rendre d_2 place d'entrée d'un arc inhibiteur pour la transition e_4. Le franchissement de la transition e_1 (création d'un produit par le producteur 1) crée un jeton dans la place d_2 qui bloque le franchissement de la transition e_4 (la transmission au consommateur 2 est bloquée).

D'autres exemples de réseaux de Petri stochastiques seront développés dans la revue de littérature de la partie 2.4. Nous consacrerons également les chapitres 3 et 4 à l'application des réseaux de Petri stochastiques pour modéliser des wish lists sur Internet (ce concept sera explicité dans le chapitre 3) et plus particulièrement l'impact de campagnes e-mailing sur ces systèmes de wish list.

2.4 LITTERATURE SUR LES RESEAUX DE PETRI :

Il existe de nombreux articles de recherche sur les réseaux de Petri. Ces articles sont dans la majorité des cas très techniques car ils se situent dans des courants de recherche en informatique ou en recherche opérationnelle. Nous limiterons volontairement notre revue de littérature aux articles qui nous ont paru intéressants par rapport aux problématiques de la gestion. On distinguera dans l'analyse ci-dessous trois types d'articles différents :

- un premier article de Meldman, en 1977, qui introduit les réseaux de Petri dans le domaine du management ;
- une série d'articles utilisant les réseaux de Petri pour modéliser des systèmes de flux (workflow systems). La littérature sur les réseaux de Petri est vaste en ce domaine mais nous ne retiendrons ici que les articles qui nous ont paru pertinents par rapport aux problématiques de la gestion ;
- plusieurs recherches modélisant des systèmes d'organisation ou des systèmes multi-agents par les réseaux de Petri. Ce champ de recherche est également très fourni dans la littérature sur les modèles de Petri. Il nous semble particulièrement intéressant car les cas d'application sont facilement transposables à des

problématiques marketing, où les agents seraient des consommateurs et/ou des entreprises.

2.4.1 Meldman (1977) : première introduction des réseaux de Petri dans le domaine du management

En 1977, Meldman du Massachusetts Institute of Technology publie un article dans Sloan Management Review présentant les réseaux de Petri comme une nouvelle technique modélisant le comportement de systèmes d'informations pour le management. Selon lui, le challenge permanent des managers est de comprendre des systèmes d'information homme-machine complexes. Cet article fut écrit durant la période de forte croissance des MIS (Management Information Systems). En effet, il était crucial à cette époque de comprendre des systèmes intégrant à la fois des activités humaines et des processus informatiques (phénomènes relativement nouveaux à cette période). Meldman choisit les réseaux de Petri car selon lui, un système d'information pour le management doit possèder quatre caractéristiques :

1. Un nombre important de sous-systèmes interactifs (un même individu peut accomplir plusieurs tâches en même temps et un ordinateur également)
2. Un comportement fortement parallèle (les activités de plusieurs individus ou machines dans une organisation peuvent être menées en même temps)
3. Une coordination asynchrone des sous systèmes (A certaines étapes du système, l'activité d'un sous système est liée à la fin d'activité d'un autre sous système, par exemple une personne ne peut procéder à l'encaissement d'un chèque si on ne lui a pas donné le reçu)
4. Un comportement alternatif : un manager peut utiliser juste son jugement pour décider de l'augmentation de salaire de son employé

D'autre part, il insiste sur le fait qu'un modèle acceptable doit être :

1. Formel et explicite : le modèle doit permettre à l'utilisateur de déduire les conséquences de différentes configurations du système, ou de différentes survenues d'évènements.
2. Fortement modulable : on doit pouvoir décomposer le systèmes en sous-systèmes avec une compréhension claire du lien entre les sous-systèmes

3. Compréhensif : le modèle doit pouvoir intégrer une logique complète du système, incluant les phases de processus informatiques et de processus humains

Meldman illustre son article par deux problématiques représentées par les réseaux de Petri. La première est un processus d'achat impliquant un vendeur et un acheteur, illustré par le réseau ci-dessous :

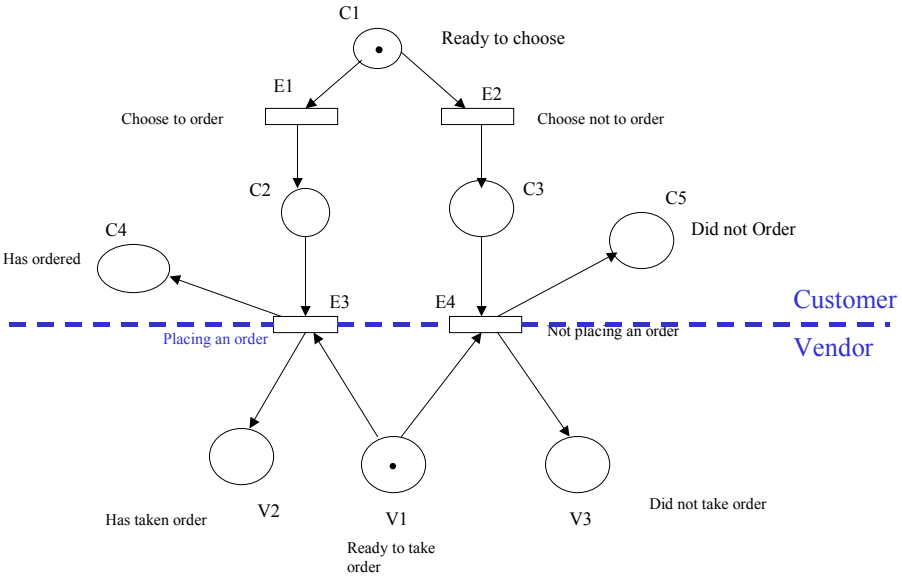


Figure 2.19 Source: Meldman, Sloan Management Review,

Dans ce système, le consommateur a le choix d'acheter (choose an order) ou de ne pas acheter (choose not to order) : le vendeur a la possibilité de prendre en charge (ou non) l'ordre d'achat. Le réseau de Petri ci-dessus représente bien cette transaction simple entre un vendeur et un acheteur.

Dans le même article Meldman modélise aussi l'influence du vendeur sur la décision d'achat d'un acheteur :

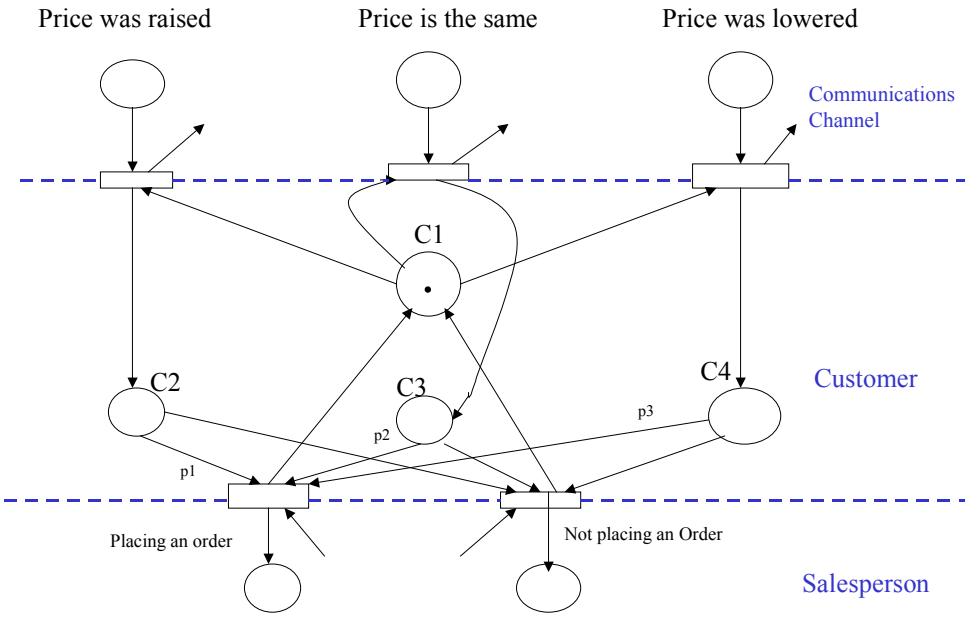


Figure 2.20 :A Petri-net representation of Vendor influence on a Buying Decision (Source: Meldman, Sloan Management Review, 1977)

La décision d'achat est modélisée comme un choix libre du consommateur. Cependant, lorsque le consommateur effectue son choix, il peut être influencé par une action du vendeur, comme par exemple ici le niveau de prix du produit. Dans le graphique ci-dessus, l'influence du vendeur est modélisée par le fait que la décision de prix du vendeur pousse le consommateur à aller dans l'un des trois états C2, C3 ou C4, chacun avec les probabilités respectives p1, p2 ou p3 pour la réalisation de l'événement « placing of an order ». Plus grande sera la différence entre ces trois probabilités, plus grand sera le degré d'influence du vendeur sur l'acheteur.

2.4.2 Les réseaux de Petri utilisés pour modéliser des systèmes de flux

Méta-analyse

Salimifard et Wright ont publié en 2001 dans European Journal of Operational Research une méta-analyse sur l'application des réseaux de Petri à la modélisation de systèmes de flux. Ils montrent que les réseaux de Petri constituent l'une des principales techniques pour mener des analyses à la fois quantitatives et qualitatives et modéliser de tels phénomènes. Ces systèmes de flux sont constitués de plusieurs activités ordonnées effectuées avec un objectif global de l'ensemble de ces activités. Celles-ci peuvent être effectuées par des ressources humaines ou informatiques au sein d'une entreprise, l'objectif global étant par exemple celui de

l'entreprise. L'automatisation et la compréhension des flux permet à l'entreprise de mieux maîtriser son activité et de s'adapter plus rapidement à un environnement concurrentiel changeant. Les systèmes de flux sont des processus au cours desquels des documents, des informations ou des tâches circulent d'un participant à un autre, l'ensemble de l'activité devant être coordonnée. Salimifard et Wright présentent les avantages suivants des réseaux de Petri pour modéliser de tels systèmes :

- les réseaux de Petri ont des fondations mathématiques bien définies ;
- le formalisme mathématique des réseaux de Petri permet de décrire et de modéliser le comportement du système ;
- ils sont simples à comprendre en raison de leur représentation graphique;
- les réseaux de Petri sont particulièrement adaptés pour modéliser et analyser les systèmes à événements discrets caractérisés par des activités parallèles et synchronisées ;
- Après avoir modélisé un système par les réseaux de Petri, il est possible de mener deux analyses, la première sur le comportement du système (lié au marquage initial), la deuxième sur ses propriétés structurelles (indépendant du marquage initial) .

Salimifard et Wright illustrent leur propos en modélisant le système de réclamation client suivante :

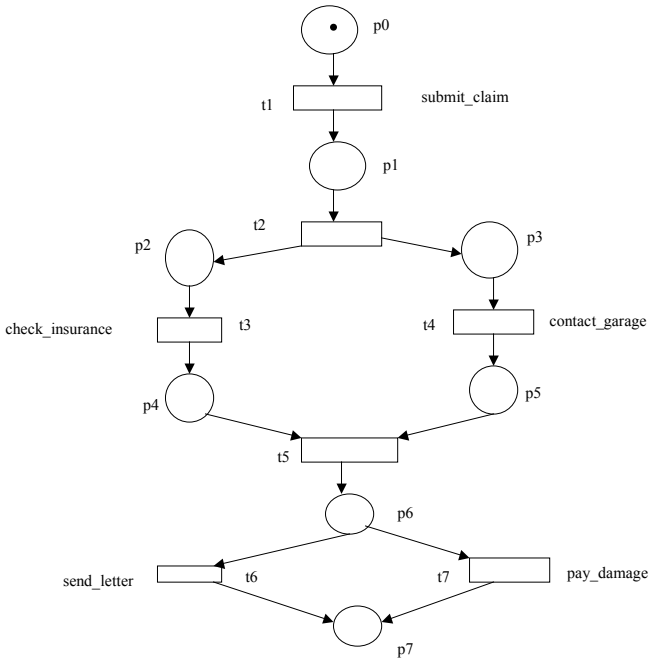


Figure 2.21 Petri net model of workflow
(Source: Salimifard and Wright, European of Operational Research, 2001)

Ce système est constitué de cinq tâches : submit_claim (le consommateur soumet sa réclamation), check_insurance (il vérifie son assurance), contact_garage (il contacte le garage), send_letter (il envoie une lettre) et pay_damage (il paie lui même les réparations).
Le système commence par la place p0 et se termine par la place p7.
Chaque processus est représenté par une transition. On fait l'hypothèse que certaines tâches sont effectuées simultanément (vérifier l'assurance et contacter le garage par exemple).
En fonction du marquage initial du système, on obtiendra le graphe d'état du réseau détaillant tous les états et les événements de ce réseau . L'analyse du graphe d'états permettra de détecter les propriétés du réseau (déceler notamment les points de blocage, très préjudiciables à la performance du système global).
Les auteurs étendent ensuite ce modèle aux réseaux de Petri temporisés, dans lesquels on peut paramétrer une durée pour chaque transition. Ils détaillent ensuite les analyses qualitatives et quantitatives du système que permettent les réseaux de Petri. Les premières consistent à détecter les points de blocage du système (grâce au graphe d'état) et de savoir si la séquence des différentes étapes est correcte, s'il n'y a pas de conflits (deux transitions suivant une place ne comportant qu'un seul jeton par exemple) dans ce système, si le système peut être validé. L'analyse quantitative consiste à mesurer des critères de performance du réseau : quelles sont les ressources utilisées ? Quelle est la durée d'une tâche avant qu'elle ne soit traitée ?
Salimifard et Wright évoquent également dans l'article les extensions des réseaux de Petri, comme les réseaux colorés (permettant d'identifier par exemple différents types de consommateurs), ou les réseaux de Petri stochastiques généralisés (GSPN), chaque transition et place présentant une probabilité de durée particulière.

Après avoir décrit l'approche de Salimifard et Wright, nous allons présenter ci-dessous deux articles utilisant les réseaux de Petri pour modéliser des systèmes de flux. Ces recherches sont des applications intéressantes pour le domaine de la gestion : l'un porte sur la modélisation de supply chains, l'autre porte sur le E-commerce.

Supply chain

En 2000, Journal of the Operational Research Society publie un article sur la modélisation de supply chains par des réseaux de Petri stochastiques généralisés. Le processus d'arrivée des commandes clients est modélisé par une loi de Poisson et les services de la supply chain par des lois exponentielles. Les auteurs comparent ensuite les performances de plusieurs systèmes

différents (en terme de planning, de contrôles, de stockage, de coûts, etc…), en modélisant le processus par des réseaux de Petri stochastiques.

Le supply chain management (SCM) est la coordination ou l'intégration d'activités de plusieurs entreprises impliquées dans la production et la livraison de produits ou services destinés à des consommateurs situés dans des lieux géographiques différents. Traditionnellement, les entreprises agissent de manière indépendante, le SCM permettant de coordonner toutes ces organisations et toutes les fonctions impliquées jusqu'au processus de livraison. La modélisation de ces systèmes complexes est cruciale pour évaluer la performance et comparer différents systèmes de supply chain.

Le système modélisé par Viswanadham et Srinivasa Raghavan est le suivant :

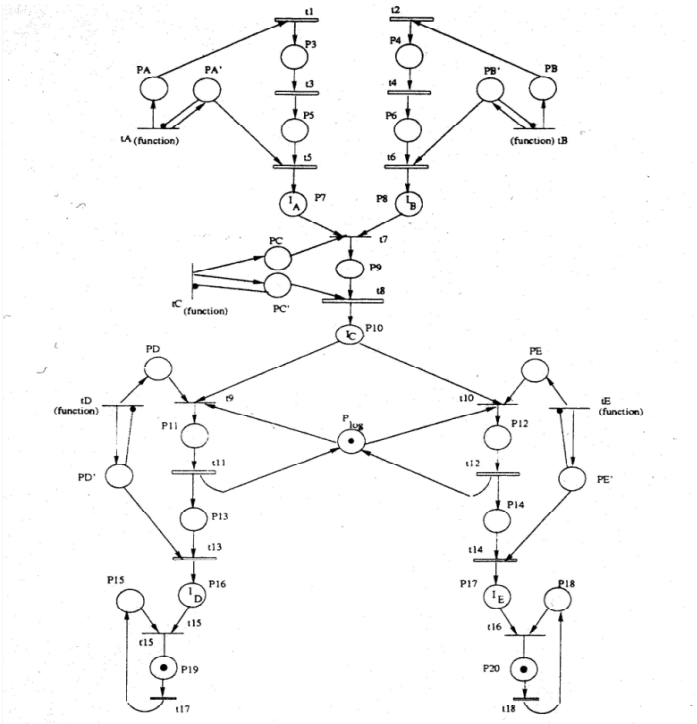


Figure 2.22 : Supply chain

Source: Viswanadham et Srinivasa Raghavan (2000)

Tableau 2.2: Interprétation des places dans le réseau de Pétri de la figure 2.22

Nom de la place	Description
PA'	Material on order to supplier of A
PB'	Material on order to supplier of B
PA	Manufacturing at supplier of A
PB	Manufacturing at supplier of B
P3	Logistics from supplier of A
P4	Logistics from supplier of B
P5	Interface between supplier A logistics and OEM (OEM:original equipment manufacturers)
P6	Interface between supplier B logistics and OEM
P7	Available inventory of A
P8	Available inventory of B
PC	Order receipt for production of C
PC'	Material on order for production of C
P9	Production of C
P10	Inventory of C available
PD	Order receipt for production of D
PD'	Material on order for production of D
PE	Order receipt for production of E
PE'	Material on order for production of E
P11	Outbound logistics of D from plant to warehouse
P12	Outbound logistics of E from plant to warehouse
Plog	Logistics carriers available
P13	Assembling of D from inventory of C
P14	Assembling of E from inventory of C
P16	Finished goods inventory of D at warehouse
P17	Finished goods inventory of E at warehouse
P15	Back order for D ready
P18	Back order for E ready
P19	Customer order for D ready
P20	Customer order for E ready

Tableau 2.2 (suite): Interprétation des transitions du réseau de Pétri de la figure 2.22

Nom de la transition	Description
tA	Start of manufacturing of A
tB	Start of manufacturing of B
t1	Processing by supplier of A
t2	Processing by supplier of B
t3	Transportation from supplier of A
t4	Transportation from supplier of B
t5	Paper work on interfaces with supplier of A
t6	Paper work on interfaces with supplier of B
tC	Trigger for production of C
t7	Manufacturer of C starts production
t8	Processing of C
tD	Trigger for assembling of D
tE	Trigger for assembling of E
t9	End of assembling of D from C
t10	End of assembling of E from C
t11	Outbount logistics of D
t12	Outbount logistics of E
t13	Assembling of D
t14	Assembling of E
t15	Customer order for D served
t16	Customer order for E served
t17	Arrival of order for D
t18	Arrival of order for E

Le détail du fonctionnement de ce réseau peut être trouvé dans Viswanadham et Srinivasa Raghavan (2000). Les auteurs détaillent et expliquent en particulier chacune des étapes du modèle complexe ci-dessus présenté. Ils s'intéressent ensuite aux mesures de performances de la supply chain, et notamment aux coûts d'un tel système. Etant donné la décomposition du système en états et transitions, ils évaluent le coût global du système, et comparent trois modèles de supply chain :

1. le modèle Make-to-Stock (MTS): les produits du consommateur final proviennent de stocks de produits finis stockés dans différents entrepôts.
2. le modèle Make-to-order (MTO) : la confirmation d'une commande d'un consommateur entraîne le déclenchement des différentes étapes du système. Chaque ordre de consommateur est unique en terme de production, logistique, etc. Il n'y a que très peu ou pas de stocks.
3. le modèle Assemble-to-order (ATO) : Une variété de produits sont assemblés à partir de composants et de sous-assemblages.

Les auteurs font l'hypothèse que la durée entre deux commandes consommateurs suit une loi exponentielle, et mesurent donc l'impact d'une modification des paramètres de la loi exponentielle sur les coûts du système. Cette simulation leur permet de comparer les performances des trois types de supply chain MTS, MTO et ATO, en fonction de différentes hypothèses de durée intercommandes.

E-commerce

En 1998, Weitz publie dans Lecture notes in Computer Science un modèle de flux fondé sur les réseaux de Petri pour modéliser le e-commerce. Il applique les réseaux de Petri pour représenter les échanges de flux d'informations dans le cadre du commerce électronique, notamment la génération de catalogues électroniques, le processus de commande en ligne, les échanges de documents, de forme HTML ou XML, qui sont les deux formes dominantes de documents standards circulant sur Internet. Weitz utilise les réseaux de Petri pour représenter le processus de commande d'un consommateur sur Internet.

Selon Weitz, le problème des modèles de flux classiques est qu'ils ne capturent pas la structure (parfois complexe) des documents échangés. Les réseaux de Petri quant à eux permettent de créer et de manipuler des documents avec des structures spécifiques qui sont détaillées dans le schéma ci-dessus, puisqu'on peut spécifier sur chaque arc du réseau de Petri le type et la structure du document qui circule dans le processus de flux.

2.4.3 Les réseaux de Petri pour modéliser des systèmes d'organisation ou des systèmes multi-agents :

Systèmes d'organisation

Ce courant de recherche nous semble intéressant à creuser car les problématiques traitées sont transposables au domaine du marketing qui nous concerne (les agents seraient des consommateurs ou des entreprises).

En 1996, un article de Tsalgatidou et Louridas en 1996 dans Simulation & Gaming présente les réseaux de Petri pour modéliser et simuler des systèmes d'organisation dynamiques. Selon les auteurs, la mission d'une organisation est la réalisation de ses objectifs grâce à l'exécution de différents processus et méthodes spécifiques à l'organisation de l'entreprise. La modélisation des processus dynamiques est essentielle pour comprendre la fonctionnalité de l'organisation, pour découvrir des problèmes existants et pour proposer et évaluer des

solutions alternatives. Les réseaux de Petri permettent, lors de la validation dynamique du système, de déterminer les points de blocage et de détecter les erreurs du processus. Ils constituent donc des outils importants pour les managers.

Systèmes multi-agents

La modélisation des systèmes multi-agents par les réseaux de Petri a été appliquée dans des sciences diverses et en particulier en biologie. Nous allons ci-dessous détailler un article de Bakam, Kordon, Le Page et Bousquet qui appliquent par exemple les réseaux de Petri afin d'étudier un système de gestion de la chasse au Cameroun. Bien que cette recherche soit loin des problématiques qui se posent en marketing, elle nous semble intéressante en terme de potentiel de transposition à notre domaine, où les agents seraient des consommateurs ou des entreprises. Généralement, les systèmes multi-agents permettent de représenter des entités autonomes en interaction, dotées de comportement et pouvant évoluer dans un environnement. L'objectif est d' analyser le système et d'en identifier les propriétés globales. Le cas de la chasse au Cameroun est spécifique : les auteurs étudient la chasse au céphalophe bleu (espèce la plus chassée). Les enquêtes terrain leur ont permis de recueillir des données sur le comportement de cet animal, les types de pièges, le volume des prises et les stratégies des chasseurs. La question posée est de savoir si les règles locales vont dans le sens d'une dégradation de la ressource ou si elles forment de facto un système viable de gestion.

Les auteurs ont pour cela :

- simulé les comportements sociaux des céphalophes (marquage des territoires, reproduction) pour étudier l'émergence de l'organisation sociale (familles)
- construit et étudié par simulations le modèle multi-agent de l'activité de chasse, défini par :
    - un espace fait d'un quadrillage de cellules qui possèdent des attributs représentant l'état du sol (eau, route, piège,…)
    - des agents animaux définis par l'âge, le sexe, la position dans l'espace
    - des agents chasseurs qui choisissent des zones de piégeage au cours des saisons

Les auteurs ont ensuite modélisé ce système par les réseaux de Petri (en effectuant notamment des hypothèses sur le déplacement des agents, la croissance de la population animale, le séquencement des différentes actions) et ont mené deux simulations : la dynamique hors

chasse (les comportements des animaux dans leur environnement naturel) ; l'activité de chasse (comportement des chasseurs qui alternent les aires de chasse).

Bakam, Le Page et Bousquet montrent ainsi par ces simulations sur les réseaux de Petri que la distribution spatiale et les stratégies des chasseurs ont un impact sur la viabilité de la ressource.

Nous allons détailler enfin un dernier article sur les systèmes multi-agents modélisés par les réseaux de Petri, article qui présente une approche là aussi transposable au domaine du marketing. il s'agit d'un papier présenté à la conférence MOCA 2002, par David Weyns et Tom Holvoet. Ces auteurs utilisent un réseau de Petri coloré pour modéliser une application multi-agent : le « Packet-World », cas d'application pour étudier le comportement social d'agents. L'une des originalités de leur démarche est de combiner des expériences et des modèles conceptuels. Un autre élément intéressant est qu'ils présentent tout d'abord un modèle de base puis le complexifient en ajoutant de nouveaux comportements sociaux des individus, et en faisant évoluer les agents et l'environnement. Pour ces auteurs un système multi-agent est une société d'agents qui vivent et travaillent ensemble. Or, vivre dans une communauté requiert un certain nombre de compétences sociales. Le « packet-world » est un monde virtuel qui comporte plusieurs paquets colorés placés au hasard dans certaines cases d'un rectangle. Les agents vivant dans ce système doivent collecter ces paquets et les porter jusqu'à une case de la même couleur que le paquet. Les agents ont une vue limitée du monde , mais peuvent partager leurs informations, coopérer pour atteindre leur objectif.

Ce système est représenté dans la figure 2.23 ci-dessous :

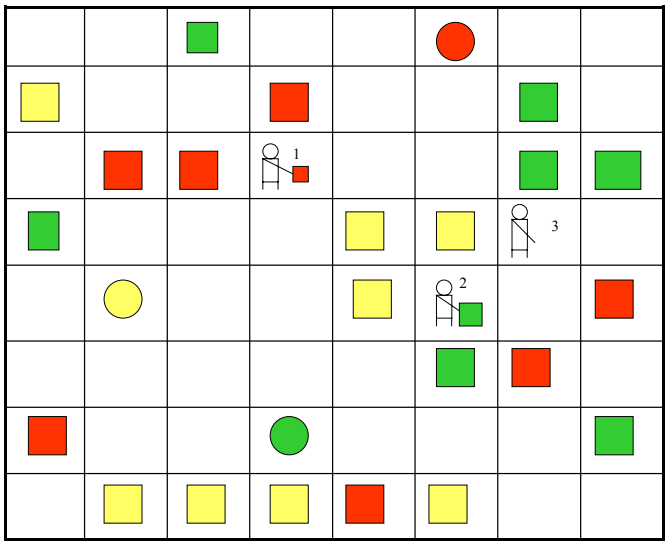


Figure 2.23 Le « packet-world » (les carrés représentent les paquets, les cercles sont les points de livraison)
Source: Weyns and Holvoet, 2002

Dans l'exemple ci-dessus, la grille contient une destination pour chaque couleur. C'est un « packet-world » de taille 8 avec 3 agents. Dans ce système, les agents peuvent interagir de différentes manières :

- un agent peut aller dans l'une des cases voisines de la case dans laquelle il se trouve,
- si un agent ne porte pas de paquets, il peut en prendre un dans l'une des cases voisines de celle dans laquelle il se trouve,
- un agent peut poser le paquet qu'il porte sur l'une des cases voisines ou bien sûr sur la case de destination de ce paquet,
- un agent peut attendre sans rien faire.

Il est important de noter que chaque agent a une vision limitée du monde. Sa vue couvre seulement une partie de l'environnement qui l'entoure (le nombre de cases qu'il voit autour de lui).

Pour modéliser la performance des agents, un compteur s'incrémente chaque fois qu'un agent avance d'une case ou bouge un paquet (soit en le prenant, soit en le posant). A chaque instant la valeur de ce compteur indique l'énergie qu'a dépensé l'agent dans sa tâche jusque-là.

L'objectif du système est d'amener les paquets jusqu'à leur destination finale avec le minimum d'énergie dépensée.

Dans le modèle de base, les agents n'interagissent pas : ils répètent simplement un cycle d'actions : regarder les paquets autour d'eux, les prendre, regarder la destination et livrer le paquet. Dans le modèle étendu, les agents peuvent interagir entre eux et coopérer en partageant l'information qu'ils possèdent. Plusieurs types de coopération peuvent être envisagées (les agents font une chaîne et se passent les paquets les uns les autres par exemple). Les auteurs choisissent une autre forme de coopération : au lieu d'explorer le monde pour trouver une cible que l'agent ne voit pas, il demande de l'information aux autres agents visibles. Si cet agent a la réponse à la question posée, il continue sa route jusqu'à la destination finale qui lui permet d'atteindre son objectif.

Weyns et Holvoet modélisent ce système par des réseaux de Petri colorés (les couleurs des jetons permettent de différencier les agents entre eux, les cases entre elles et les paquets entre eux). Pour plus de détail sur le réseau de Petri, on peut se référer à l'article de Weyns et Holvoet.

En modélisant un tel système par les réseaux de Petri, les auteurs finalement comparent (grâce à la simulation) les performances des deux systèmes décrits : le modèle de base et le modèle dans lequel les agents communiquent. Ils comparent également des tailles de packet-world

différents (nombre de cases et/ou nombre d'agents plus importants) et différentes capacités des agents (nombre de cases qu'ils peuvent voir). Ils montrent que la communication entre les agents les rend plus efficaces et que la valeur de l'information échangée entre les agents est importante.

CONCLUSION :

Nous avons dans ce chapitre 2 de la thèse introduit les modèles des réseaux de Petri. Nous avons tout d'abord présenté les éléments de base et les principales caractéristiques de ce type de modèles. Nous avons ensuite cherché à rendre plus concrète la présentation des réseaux de Petri en détaillant deux applications possibles: la modélisation de systèmes de listes de cadeaux sur Internet (wish lists) et le processus de commande d'un client, ce qui a permis de souligner et mettre en avant les riches possibilités des réseaux de Petri pour modéliser des systèmes dynamiques pouvant être parfois complexes. Après une synthèse sur les différentes familles existantes des réseaux de Petri, nous avons détaillé les spécificités de l'une de ces familles, les réseaux de Petri stochastiques (SPN), ceux-ci constituant les modèles utilisés dans les applications empiriques de la thèse (chapitre 3, 4 et 5). Enfin, nous avons terminé ce chapitre par une revue de littérature sur les réseaux de Petri, se limitant volontairement aux articles ayant trait à des problématiques proches du domaine de la gestion qui nous intéresse.

Ce chapitre visait à souligner la flexibilité et la richesse des réseaux de Petri, leur représentation graphique simple permettant de simplifier la compréhension de systèmes parfois complexes, leur robustesse mathématique et l'intérêt suscité par ces modèles dans la littérature pour traiter des problématiques liées au domaine de la gestion.

Nous allons dans les chapitres suivants utiliser les réseaux de Petri pour modéliser des systèmes multi-agents interactifs sur Internet.

Chapitre 3
Application des Réseaux de Petri à la modélisation des systèmes de wish list

3.1 INTRODUCTION :

Dans ce chapitre, notre objectif est d'effectuer une première application des réseaux de Petri dans le domaine du marketing. Nous introduisons ces modèles pour représenter de manière dynamique les comportements de visites et d'achats sur un site Web, le cas d'application empirique étant un site de wish lists. Les wish lists sont des listes de cadeaux que l'internaute peut créer pour lui même ou pour des personnes de son entourage, les cadeaux étant sélectionnés sur différents sites Internet lors de la navigation sur le Web. Le concept de wish list s'inspire de celui des listes de mariage classiques que plusieurs personnes peuvent consulter pour effectuer un cadeau. La wish list est une liste permanente de tous les produits souhaités par la personne qui la crée, et qui peut être utilisée à toute occasion de la vie quotidienne, soit pour des évènements particuliers (anniversaire, Noël, fête des mères, St-Valentin, etc…) , soit pour une autre raison (par exemple lors d'une invitation pour un dîner : les amis peuvent porter un cadeau de la wish list plutôt que plus classiquement un bouquet de fleurs). Pour modéliser ces systèmes de wish list par les réseaux de Petri, nous allons décrire et modéliser les différentes activités sur le site Web ainsi que leurs transitions avec des distributions de probabilité qui seront ensuite intégrées dans le modèle de Petri pour représenter le système global. La comparaison de la performance des réseaux de Petri avec d'autres modèles sera effectuée au chapitre 4.

Ce chapitre est construit de la manière suivante : le cadre conceptuel de cette recherche est présenté dans une première partie, suivi du cadre méthodologique utilisé (les réseaux de Petri stochastiques). Enfin, nous détaillerons le modèle de conclure sur l'ensemble de cette recherche.

3.2 CADRE CONCEPTUEL

Notre objectif est d'appliquer les réseaux de Petri stochastiques (présentés dans le chapitre 2 de la thèse) pour modéliser les systèmes de wish list. La littérature en marketing concernant la modélisation de phénomènes sur Internet, que nous avons détaillée dans la partie 1.6.2, ne comporte, selon l'état actuel de nos connaissances, aucun article concernant la modélisation de wish lists. Plus globalement, peu de recherches en marketing concernant Internet

développent des modèles intégrés permettant de représenter plusieurs activités séquentielles généralement menées par l'internaute lors de sa navigation sur un site Web. Nous allons dans cette partie re-citer les principaux articles de modélisation du comportement de l'internaute issus de la littérature en Marketing afin de préciser le cadre conceptuel que nous utilisons pour notre application. Ces articles ayant été détaillés dans le chapitre 1 de la thèse, nous nous consacrerons plus particulièrement à les positionner par rapport au modèle et à l'application que nous proposons dans ce chapitre 3.

Comme nous l'avons détaillé dans la partie 1.6.2 du chapitre 1, plusieurs articles en Marketing visent à modéliser le comportement de visite et d'achat sur site Web. Ces articles sont particulièrement récents, ce qui démontre l'important courant de recherche actuel en Marketing sur la modélisation du comportement de l'internaute sur un site Web. Les quatre principales recherches dont nous nous sommes inspirées pour notre application empirique sont les suivants. Le premier article est celui de Sismeiro et Bucklin (2004) qui proposent un modèle de prédiction de l'achat sur Internet, basé sur les différentes tâches que l'internaute doit accomplir : (configuration du produit, entrée de toutes les informations personnelles, confirmation de commande, fourniture du numéro de carte bancaire et validation de la transaction). En utilisant des données de navigation de l'internaute, ils décomposent le processus de comportement d'achat sur le Web en une séquence de différentes tâches. Notre recherche poursuit ce courant et décompose également le comportement de l'internaute en différentes actions. En revanche, Sismeiro et Bucklin utilisent dans leur modèle des chaînes de Markov décriant des processus stationnaires. Le modèle que nous utilisons (réseau de Petri stochastique) permet de modéliser aussi des processus non stationnaires.

Telang et al. (2004) proposent une extension des modèles stochastiques concernant la durée inter-achat des internautes. Ils utilisent des modèles proportionnels mixtes de hasard pour prédire les visites sur un moteur de recherche. Nous utilisons également des modèles stochastiques pour modéliser les durées inter-vistes ou inter-achats. Cependant, Telang et al. (2004) n'intègrent pas l'impact des visites sur l'achat, comme nous nous proposons de le faire.

Moe et Fader (2004b) lient l'achat et le comportement de visite de l'internaute en développant un modèle de comportement dynamique permettant de prédire les achats sur un site en fonction des précédentes visites. A la fin de leur article, les auteurs proposent comme piste de future recherche de développer un modèle intégré combinant des modèles stochastiques de visites et de comportements d'achat. Nous développons un tel modèle dans ce chapitre de la thèse, grâce à l'application des réseaux de Petri stochastiques à la modélisation de systèmes

de wish lists. . Dans une autre recherche (2004a), Moe et Fader proposent un modèle basé sur des données individuelles afin d'étudier l'évolution du comportement de visite de l'internaute sur un site Web. Ils introduisent un modèle non stationnaire en ajoutant un facteur multiplicatif au paramètre lambda de la distribution exponentielle modélisant la durée inter-visite. Ils montrent que plus l'internaute effectue des visites sur le site (il devient alors plus expérimenté), plus sa propension à acheter est importante. Ce type de modèle non stationnaire décrivant la durée inter-visite pourrait tout a fait être intégré dans un modèle de Petri. Plusieurs autres modèles sur le comportement des internautes ayant été détaillés dans le chapitre 1, nous ne les re-présentons donc pas ici.

Le premier constat qu'on peut tirer de cette littérature en marketing est qu'aucune de ces recherches n'étudie le comportement des internautes sur un site Web par une approche multi-agents, alors que plusieurs internautes effectuent des actions différentes au même moment lorsqu'on analyse un site Web et que de nombreux phénomènes constituent des systèmes multi-agents (communautés virtuelles, plusieurs internautes naviguant sur un site Web, etc…). D'autre part, les modèles ci-dessus présentés présentent plusieurs limites, comme tout d'abord leur complexité. Tout d'abord, la taille importante des bases de données disponibles sur le Web (Moe et Fader, 2004b) rend l'utilisation de ces modèles difficile et leur implantation en entreprise peu probable. La plupart des recherches ci-dessus n'intègrent pas non plus tous les éléments de navigation de l'internaute, limite souvent fournie à la fin des différentes recherches ci-dessus publiées. Enfin, aucune de ces recherches ne portent sur l'étude du phénomène de wish lists, concept marketing pourtant innovant sur le Web.

Nous allons donc dans ce chapitre appliquer les modèles des réseaux de Petri pour représenter un système de wish list sur Internet et montrer que ce type de modèle permet de modéliser des systèmes complexes en intégrant de nombreux paramètres du comportement de plusieurs internautes sur un site Web dans un système global.

Modéliser le comportement sur le Web dans une perspective globale est important pour plusieurs raisons. Rappelons (cf chapitre 1) qu'une étude d'avril 2000 du Boston Consulting Group et de shop.org (Betts, 2001), les sites Web reçoivent des millions de visiteurs mais seulement 3% d'entre eux achètent, que les visites d'un site Web ne sont pas de bons éléments prédicteurs des intentions d'achat, et que les modèles à un seul stade (de type probit ou logit) qui attribuent un impact direct des visites sur le comportement d'achat ne sont pas adaptés pour analyser le Web, comme le soulignent Sismeiro et Bucklin (2004).

Les caractéristiques spécifiques du comportement de visite et d'achat sur le Web (comme la simultanéité ou le parallélisme des actions de plusieurs internautes sur un site) nécessitent

donc des méthodes flexibles et adaptées pour modéliser les actions des différents acteurs du système considéré. Les modèles classiques de causalité, en particulier les modèles single-stage, qui explorent une partie d'un système complexe, sont limités pour analyser l'aspect dynamique des comportements de visites et d'achat sur le Web. En appliquant les réseaux de Petri à l'application empirique du système de wish lists, notre objectif est de montrer que les réseaux de Petri peuvent analyser des systèmes complexes, notamment constitués d'actions parallèles, séquentielles, parfois synchronisées, et que ces modèles sont particulièrement adaptés à l'étude des phénomènes sur Internet.

3.4 CADRE METHODOLOGIQUE DES RESEAUX DE PETRI STOCHASTIQUES POUR MODELISER UN SYSTEME DE WISH LIST

3.4.1 Le concept de wish list :

Cette recherche a été menée en collaboration avec la société MilleMercis. Opérateur de programmes de Marketing opérationnel on-line, MilleMercis conçoit et met en œuvre des opérations de collecte, de requalification, d'analyse et d'exploitation de bases de données. Spécialiste des bases de données comportementales on-line, MilleMercis est un acteur majeur de marketing promotionnel on line, d'emailing et de promotions on line en France. Fondée en l'an 2000 par deux chercheurs en Marketing de l'Université Paris-Dauphine, cette société est leader en France sur le marché de la liste de cadeaux. Elle gère aujourd'hui 500 000 membres ayant crée leur liste de souhaits sur le site www.MilleMercis.com, et plus de 2,5 millions de produits présents sur ces listes. La société Millemercis travaille pour plusieurs clients prestigieux, dont la SNCF, Cegetel, Club-internet , BNP-Paribas, Rue-du-commerce.

Le principe de la liste de cadeaux (ou wish list) est le suivant : MilleMercis permet à tout internaute de créer la liste des cadeaux qui lui feraient plaisir quelle qu'en soit l'occasion (anniversaire, fête des mères, naissance, pot de départ, mariage, Noël, …). Ces internautes ayant créé leur liste seront dénommés les agents A. Ils peuvent eux-même acheter des produits sur leur propre liste de cadeaux (qui est dans ce cas une liste d'achats souhaités). Les proches (dénommés agent B dans notre recherche) peuvent ainsi venir consulter le site et savoir quoi offrir à la personne qui a crée sa liste.

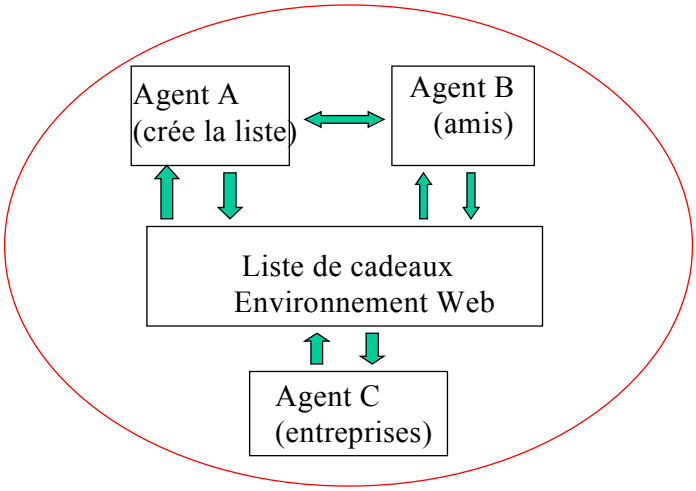
Les cadeaux de la liste peuvent être choisis via trois sources :

- les sites partenaires de MilleMercis (en cliquant sur le bouton « Faites-vous offrir ce produit »)

- les pages cadeaux du site de MilleMercis (recherche simplifiée par profil, par catégorie de produit ou par style de vie)
- le « facilitateur » qui permet de récupérer sur tous les sites Web des envies de cadeaux

Comme le montre le figure 3.1, le système d'une wish list est un système multi-agents dans lequel un agent A, des agents B et des entreprises interagissent via l'environnement Web du site www.MilleMercis.com.

Figure 3.1: SYSTEME MULTI-AGENTS DE LA WISH LIST



Les proches peuvent donc consulter les listes de cadeaux sur le site MilleMercis et aller les acheter d'un simple clic sur le site sélectionné. La liste peut être en accès libre ou restreint à certains amis (liste à accès privé).

La wish list, inspirée des listes traditionnelles de mariage off line, constitue sur le Web un concept marketing innovant et intéressant. De nombreux sites, comme www.amazon.com par exemple, fournissent actuellement un tel service pour permettre à leurs consommateurs de recevoir des cadeaux de leur choix et pour acquérir de nouveaux consommateurs enregistrés via la wish list. L'avantage majeur d'une wish list sur le Web est sa praticité d'utilisation. En quelques clicks, un internaute peut construire sa liste de cadeaux avec ses produits favoris provenant d'un site Web sur lequel il a navigué, et à la différence de traditionnelles listes (comme les listes de mariage), ses amis ou connaissances peuvent consulter facilement sa liste et lui offrir un cadeau via le Web, sans avoir à se déplacer dans une boutique. Le deuxième

avantage d'un système de liste de cadeaux est que l'utilisateur crée sa propre liste et fournit également les potentiel acheteurs (ses contacts), entraînant le fait que la société gérant les wish lists n'a pas à constituer une force de vente particulière, le concepteur de la liste faisant office également de vendeur. Enfin, le concept de wish list est particulièrement attractif en marketing puisque la liste de souhait capture les produits souhaités par l'internaute sur tout type de catégorie de produits, ce qui renforce l'efficacité d'actions de marketing direct et de promotions ciblées. En effet, les traditionnelles bases de données pour cibler les consommateurs portent sur une seule catégorie de produit, et non plusieurs. Le processus de wish list crée par Millemercis apporte plus de valeur ajoutée que des services de wish lists standard car il permet de construire une base données de produits souhaités, issus de différents sites Web, avec une information détaillée au niveau du produit. Du point de vue de la modélisation, le système de wish list constitue un site simple pour le processus de visite et d'achat : connection sur le site, sélection d'items, achat. Comparé à d'autres site de commerce en ligne, le processus de wish list est par conséquent simple et minimise la complexité de l'analyse du comportement d'achat en le considérant comme une conséquence d'une visite sur le site Web.

3.4.2 Les données utilisées:

Nous avons utilisé pour cette recherche des données fournies par la société MilleMercis, dont l'une des activités est de générer des ventes sur des sites de e-commerce via son système de wish lists. Les données fournies par MilleMercis dans le cadre de notre recherche sont constituées de 1379 listes sélectionnées par tirage aléatoire. Ces listes (toujours actives) ont été créées entre le 1/1/2000 et le 31/12/2002.

Dans un premier temps, nous avons sélectionné un sous-échantillon plus faible de ces listes, en raison des 2 éléments suivants :

1. nous cherchons à valider le modèle sur un nombre de données « acceptables » en terme de capacité informatique (les 1379 listes représentent plus de 500 000 connexions d'internautes sur la liste). Le modèle pourra être ensuite appliqué aux autres cas afin de voir s'il fonctionne également sur d'autres données (ceci augmente la validité externe de l'analyse)
2. nous avons sélectionné 207 listes particulièrement actives parmi les 1379 listes
   Cette sélection s'est effectuée sur la base des informations sur les agents B : les 207 listes sélectionnées sont des listes sur lesquelles des agents B sont enregistrés dans la

base mille mercis comme possesseur eux même d'une wish list. Dans ce cas, mille mercis récupère les données de qualification de ces agents B (critères socio-démographiques par exemple). Dans les cas où les agents B sont non possesseurs d'une liste, Millemercis ne récupère aucun critère de qualification de ces agents.

Les 207 listes sélectionnées correspondent donc à des listes particulièrement actives en terme de connexion (62 connexions de l'agents A en moyenne par liste sur les 207 versus 18,8 connexions de l'agent A sur les 1172 autres listes). Ces 207 listes ont été créées par 135 agent A , chacun de ces agents ayant en moyenne crée 1,5 listes.

Notre base de donnée comprend 22 450 connexions au site Millemercis.

Plusieurs statistiques descriptives sont fournies dans le tableau 3.1:

Table 3.1: Statistiques sur les wish lists

	Min.	Max.	Moyenne	Ecart-type
Age	8.09	77.53	30.33	10.36
Nbre de listes créées	1	15	1.53	1.48
Nbre de connexions (agent A ou B)	3	13235	278.63	1372.01
Nbre d'items achetés	0	287	6.04	29.66

Les agents A sont plutôt jeunes (30 ans en moyenne), créent peu de listes (1,5 en moyenne) et effectuent en moyenne 8 achats pour 278 connexions sur la période étudiée.

L'ensemble de l'analyse qui suit porte donc sur les 207 wish lists sélectionnées de la manière ci-dessus décrite.

3.4.3 Modèle de Petri appliqué aux wish lists

Comme nous l'avons déjà mentionné, l'objectif de notre recherche est de modéliser les systèmes de wish lists par les réseaux de Petri stochastiques.

La première étape consiste donc à décrire le système par un réseau de Petri stochastique et à le représenter graphiquement sur un logiciel dédié à ces modèles.

Plusieurs logiciels existent et peuvent être utilisés pour modéliser un phénomène par des réseaux de Petri stochastiques. La liste de ces logiciels est disponible dans le site dédié au réseaux de Petri (www.daimi.aau.dk/PetriNets). Nous avons utilisé pour cette recherche le logiciel SPNP, développé par le CACC (Center for Advanced Computing and Communication) du département « Electrical and Computer Engineering » de Duke University. Ce logiciel a été crée sous la direction du professeur Dr Kishor S. Trivedi, qui a

publié de nombreux articles sur les réseaux de Petri stochastiques. Le logiciel SPNP a été installé dans plus de 250 sites (entreprises et universités) et utilisé pour de nombreuses recherches et applications. Ces caractéristiques ont été des facteurs clés dans notre choix de logiciel. De plus, SPNP est opérationnel sous Windows, ce qui constitue un avantage important par rapport aux logiciels déjà présentés dans le chapitre 1 comme Design CPN , qui d'une part nécessite un système Linux et d'autre part ne traite pas les réseaux de Petri stochastiques.

Une wish list peut être représentée graphiquement grâce au logiciel SPNP par le réseau de Petri stochastique suivant, permettant de distinguer les différentes étapes suivantes :

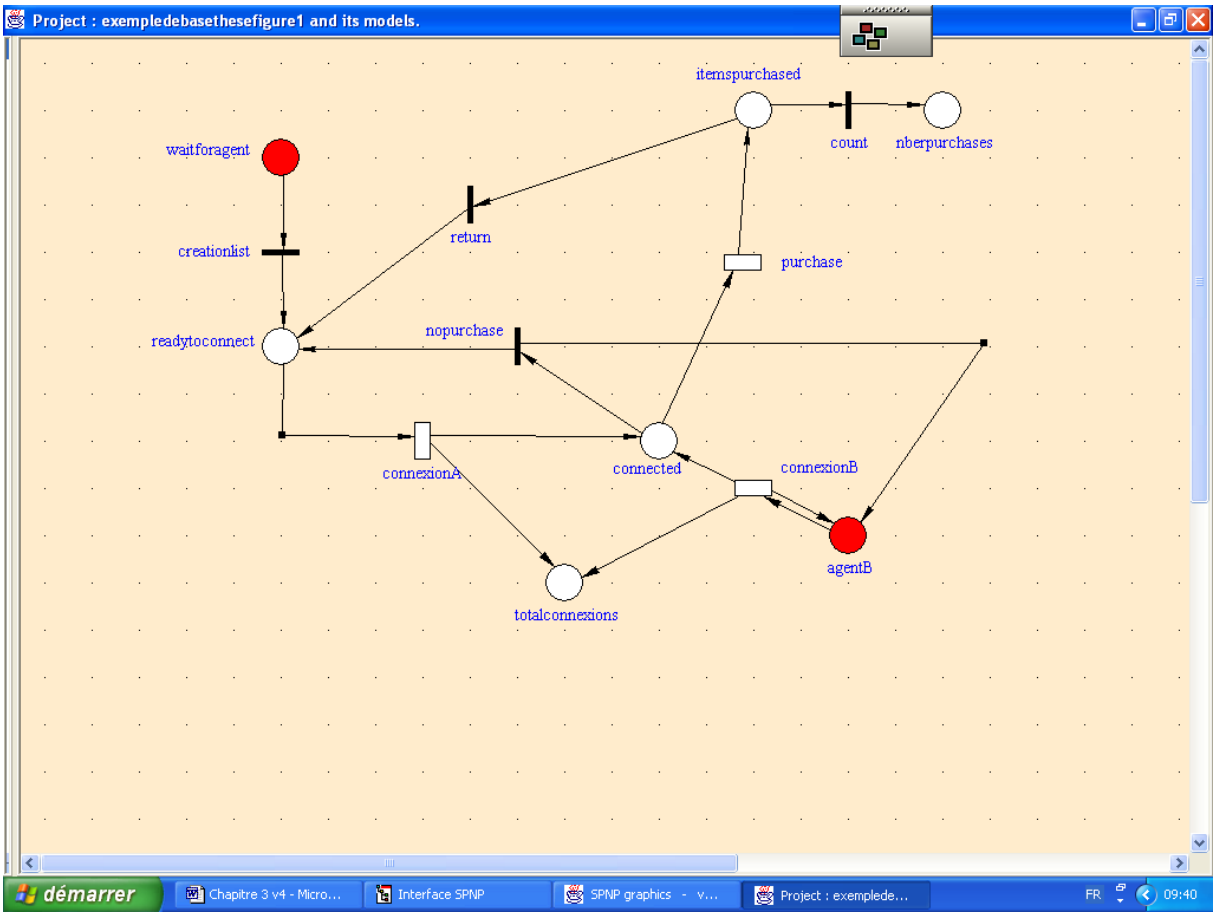


Figure 3.2 : Wish List représentée par un réseau de Petri (modèle de base)

L'interprétation de la figure 3.2 est la suivante : les agents A se trouvent à l'état initial du système dans la place *waitforagent*. Ces agents créent ensuite leur liste de cadeaux dans la transition *creationlist*. La place *readytoconnect* indique que l'état du système où la liste a été créée et l'internaute est prêt à se connecter. Lorsque l'agent A se connecte sur le site

www.Millemercis.com, la transition *connexionA* est franchie (selon une certaine loi de probabilité que nous définirons ultérieurement) et, une fois connecté (place *connected*), il achète (transition *purchase*, dont le franchissement suit une certaine loi de probabilité que nous définirons ultérieurement). Si l'agent A n'achète pas d'item lors de la connexion, la transition *nopurchase* est franchie et un jeton passe dans la place *readytoconnect*, état du système où l'agent A peut à nouveau se reconnecter (cet état peut durer plusieurs mois parfois). De même les agents B, initialement dans la place *agentB*, peuvent se connecter (transition *connexionB*) puis, après chaque connexion, acheter selon une probabilité d'achat (définie ultérieurement). Si l'agent B n'achète pas, le jeton passe par la transition *nopurchase* et revient à l'état initial de l'agent B (place *agentB*) qui peut se reconnecter sur le site.
Les items achetés sont comptabilisés dans la place *nberpurchases*, et le nombre de connexions totales (agent A + agent B) sera le nombre de jetons de la place *totalconnexions*.
Les places *waitforagent* et *agentB* sont colorées en rouge car elles contiennent à l'état initial au moins un jeton.
Ce modèle est un premier modèle que nos complexifierons par la suite. Il montre déjà qu'un réseau de Petri intègre les différentes étapes d'un processus de wish list et qu'il permet une représentation graphique simple du phénomène étudié. Comme le montre la figure 3.2, un réseau de Petri stochastique est un modèle intégré comportant plusieurs blocs décrivant des évènements modélisés individuellement: la transition *connexionA* est franchie selon une loi de probabilité que nous devons définir (loi de la variable durée entre deux visites de l'agent A), de même pour la transition *connexionB* (loi de la variable durée entre deux connexions de l'agent B). Le comportement d'achat (transition *purchase*) doit aussi être modélisé. Les lois seront paramétrées ensuite dans le réseau de Petri global.

3.4 DETAIL DU MODELE

L'objectif de cette partie est de précisément déterminer les lois stochastiques des évènements arrivant sur le système des wish lists. Ceci revient à étudier les lois de probabilités des transitions *connexionA*, *connexionB, purchase* de la figure 3.2 qui suivent des processus stochastiques (liés à la durée entre deux évènements de la même transition). En revanche, les transitions *creationlist, nopurchase, return, count* sont immédiates (pas de durée). Nous allons donc tout d'abord modéliser la manière dont les agents A se connectent sur le site de leur wish list (loi de la transition *connexionA*), puis nous modéliserons la manière dont les agents B se connectent sur le site pour visualiser la wish list de l'agent A qu'il connaît (loi de la transition

connexionB), puis nous étudierons le comportement d'achat des agents A et B (loi de la transition *purchase*).

3.4.1 Loi de durée interconnexion des agents A :

L'objectif ici est de déterminer la loi de la variable durée d'interconnexion (temps entre 2 connexions) pour chaque agent A, afin ensuite de paramétrer cette loi dans le modèle des Réseaux de Petri présenté précédemment. Notre analyse porte sur 12 839 connexions effectuées par les agents A, chiffre important confortant la validité de nos résultats.
Les données globales (sur l'ensemble des 207 listes) des fréquences des durées d'interconnexion en heure et en jour sont fournies dans le tableau ci-dessous :

Statistics

		delta connexion en heure	delta connexion en jours
N	Valid	12839	12839
	Missing	135	135
Mean		145,8778	6,0782
Median		5,3506	,2229
Mode		,00	,00
Std. Deviation		619,36481	25,80687
Variance		383612,77	665,99439
Minimum		,00	,00
Maximum		14229,97	592,92
Percentiles	10	,1256	,0052
	20	,4394	,0183
	30	,9650	,0402
	40	2,1500	,0896
	50	5,3506	,2229
	60	17,5381	,7308
	70	25,7392	1,0725
	80	70,8911	2,9538
	90	227,1058	9,4627

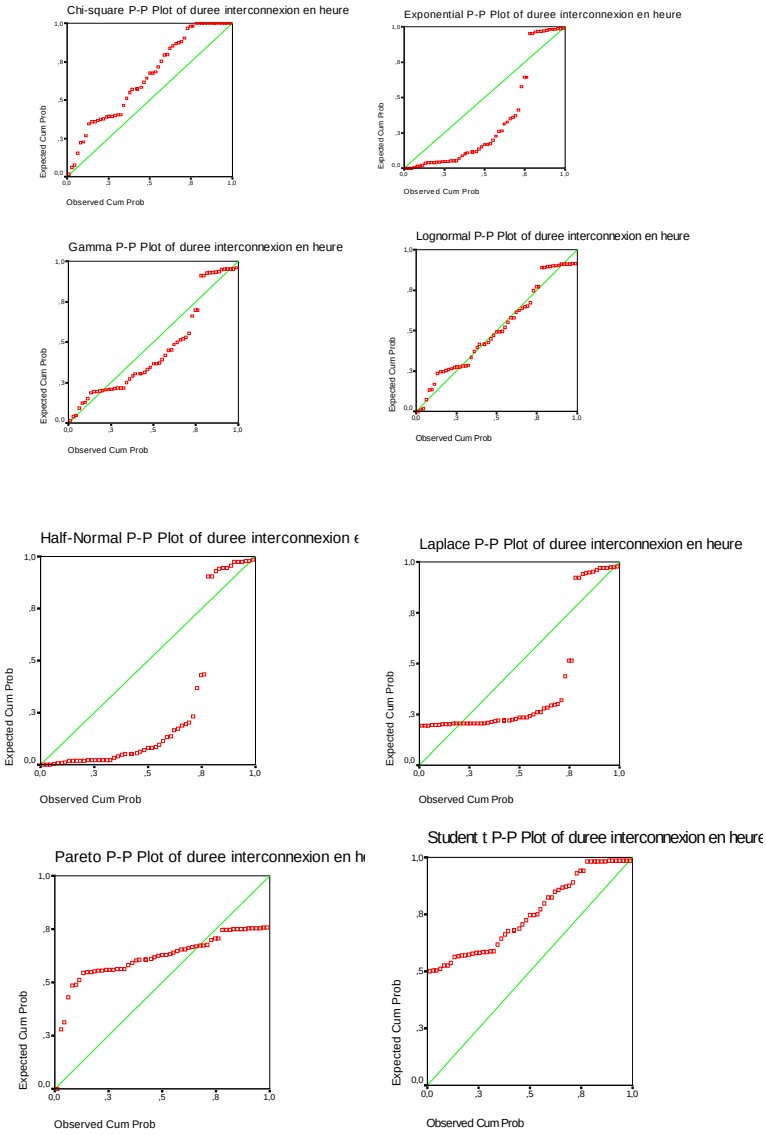
Tableau 3.2 : Statistiques descriptives durée interconnexion agent A

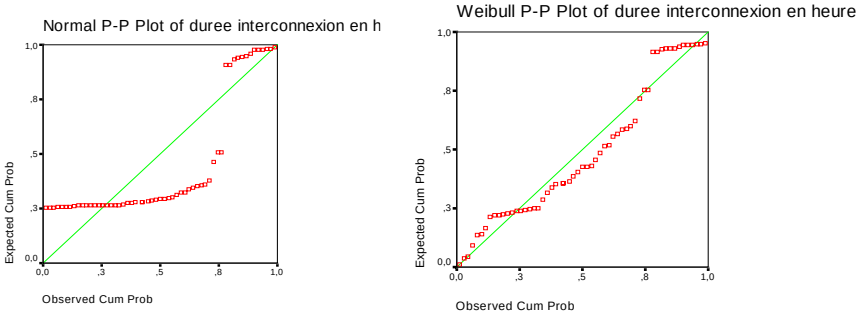
En moyenne, la durée entre 2 connexions de l'agent A est de 6 jours, soit 145 heures, avec une variance des données très importante. On constate que 70% des durées interconnexions s'effectuent en moins d'une journée (ou plus exactement de 25 heures) et la médiane des durées interconnexions est de 5 heures. L'histogramme de la répartition des durées d'interconnexion en jours est le suivant :

L'analyse globale nous ayant permis d'avoir une première vision du phénomène des connexions des agents A, nous avons ensuite cherché à modéliser la loi des durées interconnexions pour chacun des 135 agents A.

Une première analyse a consisté à étudier les PP Plots de différentes lois appliquées à la variable durée interconnexion de A. A titre d'exemple, la figure 3.3 ci-dessous montre la durée interconnexion de l'agent 130691 et permet de comparer différentes lois.

FIGURE 3.3: durée interconnexion agent A 130691





On constate sur ce cas particulier que la loi exponentielle ne donne pas une bonne représentation des données mais que trois lois donnent une représentation proche du phénomène réel : la loi Log-normale, la loi Gamma et la loi de Weibull . La deuxième justification du rejet de la loi exponentielle est qu'elle ne tient pas compte des caractéristiques de répétition des visites (Moe & fader, 2004a).

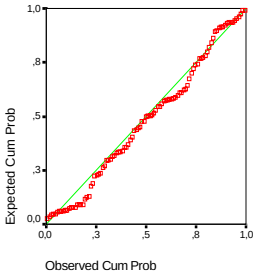
Cette démarche de comparaison des lois a été menée sur l'ensemble des 135 agents A et des résultats similaires ont été trouvés. Par conséquent, nous nous sommes concentrés sur la modélisation des durées interconnexions de l'agent A par l'une de ces trois lois (Log-normale, Gamma, Weibull).

LOI GAMMA :

Une modélisation par la loi gamma a été ensuite menée. Sur les 43 cas où le nombre d'observations est supérieur à 40 observations, les résultats sont les suivants (l'ensemble des cas est fourni en annexe 2) :

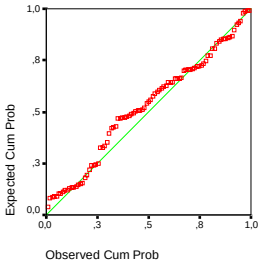
Gamma P-P Plot of duree interconnexion en heure

LOGIN: akwell1



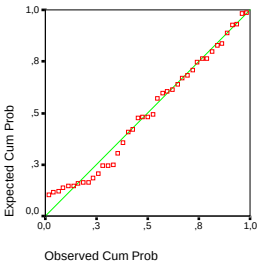
Gamma P-P Plot of duree interconnexion en heure

LOGIN: AlexO



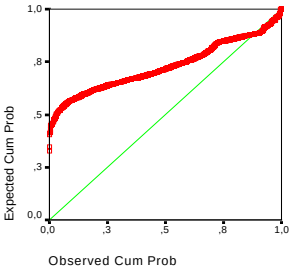
Gamma P-P Plot of duree interconnexion en heure

LOGIN: amibio



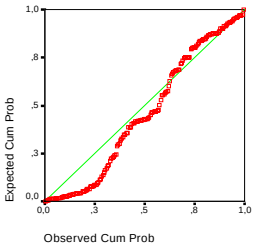
Gamma P-P Plot of duree interconnexion en heu

LOGIN: AnnA



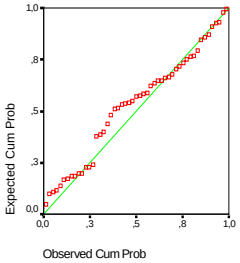
Gamma P-P Plot of duree interconnexion en heure

LOGIN: bgougay



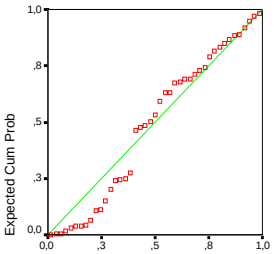
Gamma P-P Plot of duree interconnexion en heure

LOGIN: bpdl



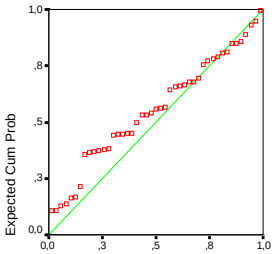
Gamma P-P Plot of duree interconnexion e

LOGIN: bschunke



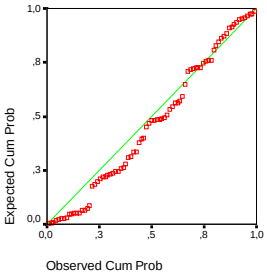
Gamma P-P Plot of duree interconnexion en

LOGIN: carolala



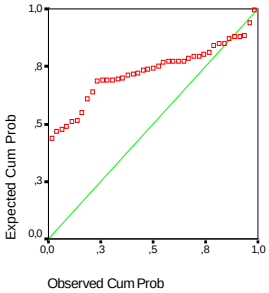
Gamma P-P Plot of duree interconnexion en heur

LOGIN: Delatte



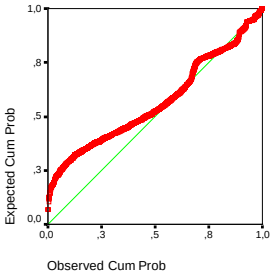
Gamma P-P Plot of duree interconnexion en

LOGIN: didine77



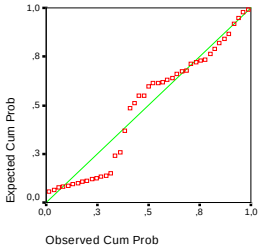
Gamma P-P Plot of duree interconnexion en he

LOGIN: djoule



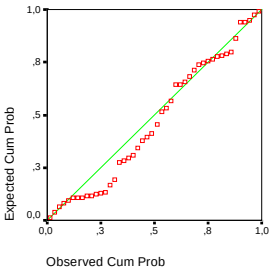
Gamma P-P Plot of duree interconnexion er

LOGIN: eb95



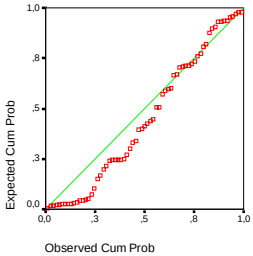
Gamma P-P Plot of duree interconnexion en heure

LOGIN: fallouis



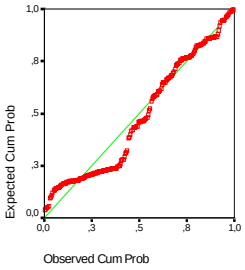
Gamma P-P Plot of duree interconnexion en heure

LOGIN: gallouin



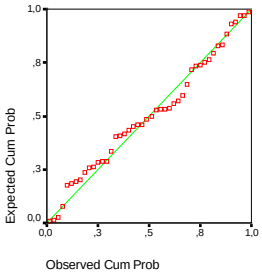
Gamma P-P Plot of duree interconnexion en heure

LOGIN: guix



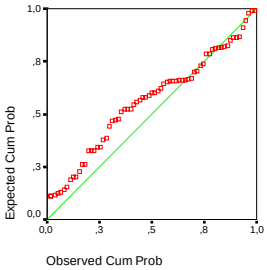
Gamma P-P Plot of duree interconnexion en heure

LOGIN: gzing



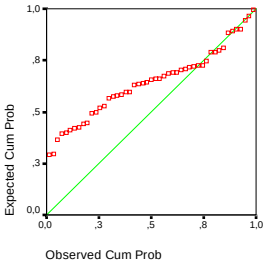
Gamma P-P Plot of duree interconnexion en heur

LOGIN: hedi.dc



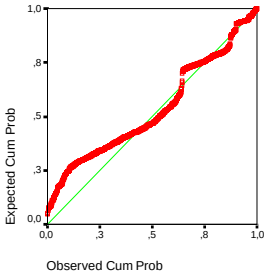
Gamma P-P Plot of duree interconnexion en heur

LOGIN: HERPIN



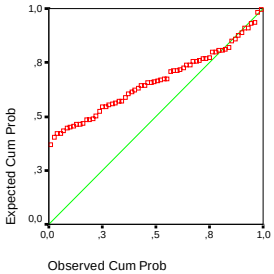
Gamma P-P Plot of duree interconnexion e

LOGIN: jeannot3



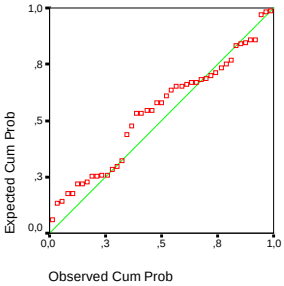
Gamma P-P Plot of duree interconnexion en heure

LOGIN: juboubou



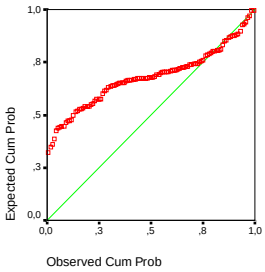
Gamma P-P Plot of duree interconnexion en heu

LOGIN: juju75



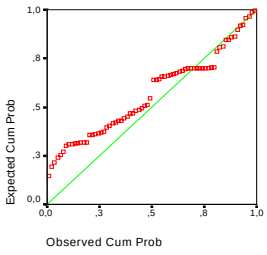
Gamma P-P Plot of duree interconnexion en he

LOGIN: kabotine



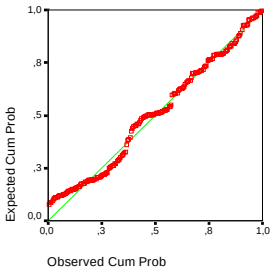
Gamma P-P Plot of duree interconnexion en heure

LOGIN: Kathrin



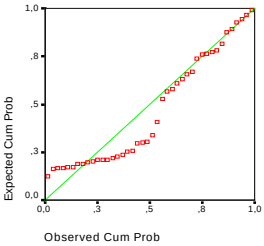
Gamma P-P Plot of duree interconnexion en he

LOGIN: Kifran



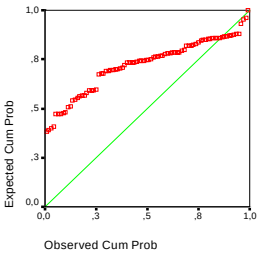
Gamma P-P Plot of duree interconnexion en l

LOGIN: levieux



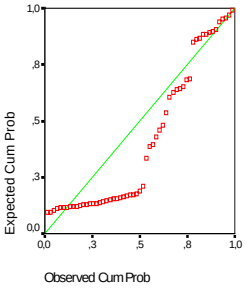
Gamma P-P Plot of duree interconnexion en heure

LOGIN: Loquet



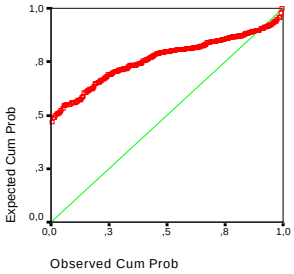
Gamma P-P Plot of duree interconnexion en

LOGIN: m@hy



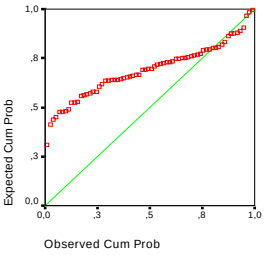
Gamma P-P Plot of duree interconnexion en he

LOGIN: mperrot



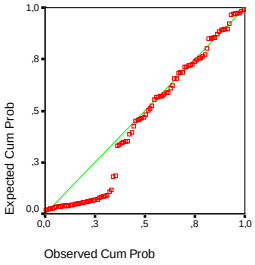
Gamma P-P Plot of duree interconnexion en heure

LOGIN: Naemesis



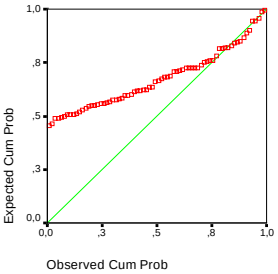
Gamma P-P Plot of duree interconnexion en heure

LOGIN: nhai07



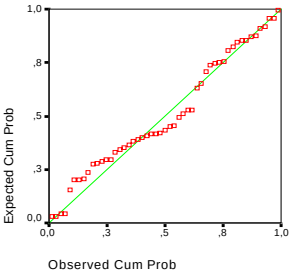
Gamma P-P Plot of duree interconnexion en hei

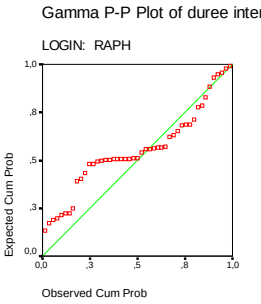
LOGIN: philia

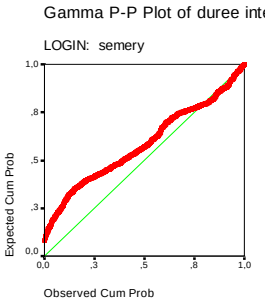


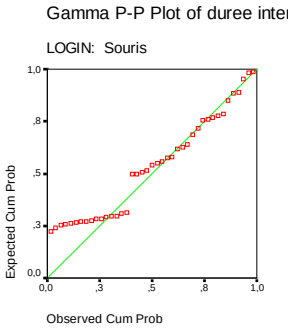
Gamma P-P Plot of duree interconnexion e

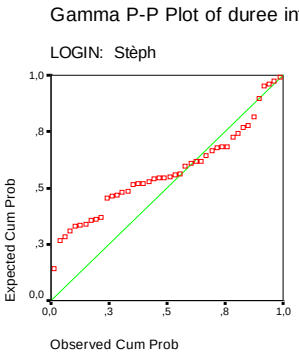
LOGIN: prozaiik

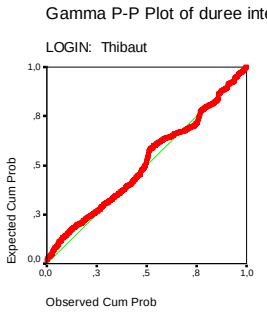


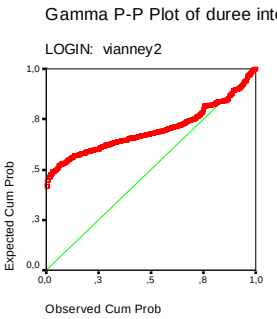


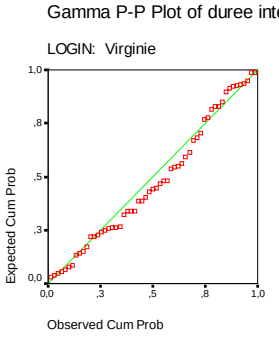










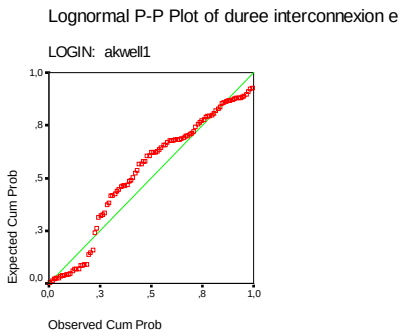


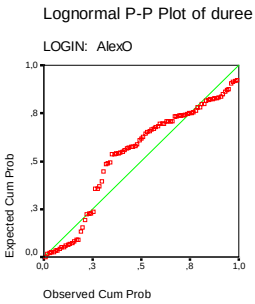
On constate sur ces graphiques que la loi Gamma ne donnent pas de bons résultats sur 12 cas sur les 43 présentés ci-dessus : Anna, didine77, djoule, Herpin, juboubou, kabotine, mperrot, Loquet, Naemesis, philia, m@hy, vianney2. Dans la majorité de ces 12 cas (tous sauf m@hy), le modèle surestime la durée interconnexion. Nous analyserons plus précisément ces cas ultérieurement. Sur les 21 autres cas, on constate que la loi Gamma est proche des données

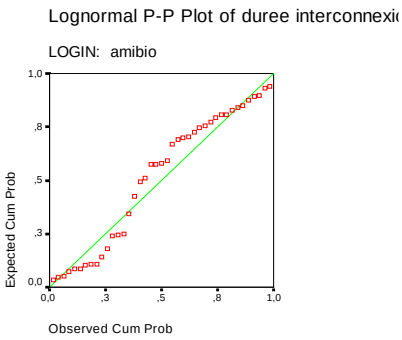
réelles. Toutefois, en raison des 12 cas précédents pour lesquels la loi Gamma représente mal les données, nous ne retenons pas pour l'instant cette loi sur la seule analyse des PP plots.

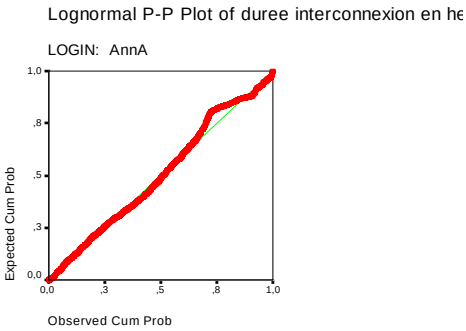
LOI LOG-NORMALE :

Une approche par la loi log-normale fournit les résultats suivant sur les 43 cas (l'ensemble des cas est en annexe 2)



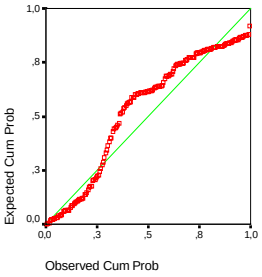






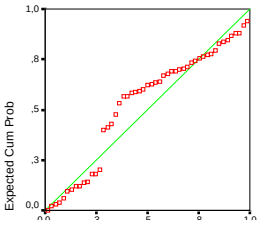
Lognormal P-P Plot of duree interconnexion

LOGIN: bgougay



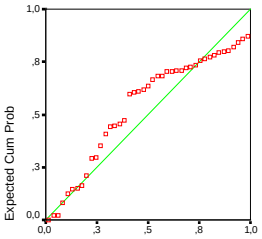
Lognormal P-P Plot of duree interconnexion en heure

LOGIN: bpdl



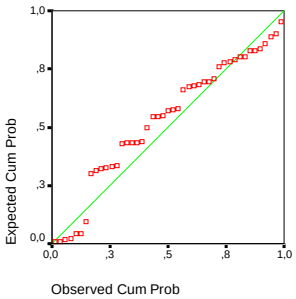
Lognormal P-P Plot of duree interconnexi

LOGIN: bschunke



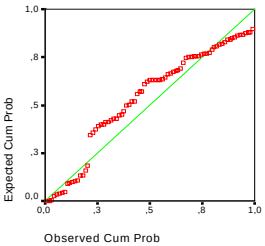
Lognormal P-P Plot of duree interconnexion en he

LOGIN: carolala



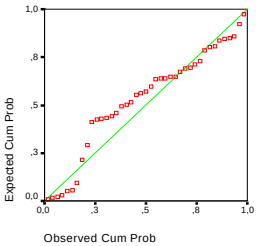
Lognormal P-P Plot of duree interconnexion e

LOGIN: Delatte



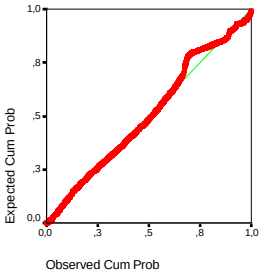
Lognormal P-P Plot of duree interconnexion en heure

LOGIN: didine77



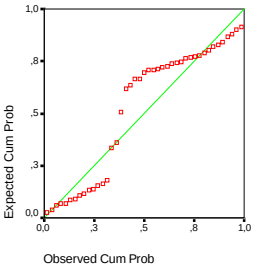
Lognormal P-P Plot of duree intercon

LOGIN: djoule



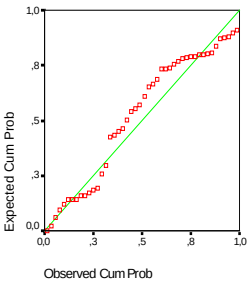
Lognormal P-P Plot of duree interconnexion en heure

LOGIN: eb95



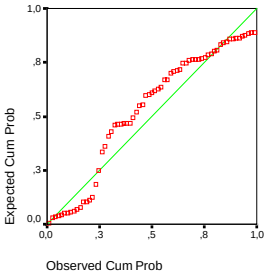
Lognormal P-P Plot of duree interconnexion en heu

LOGIN: fallouis



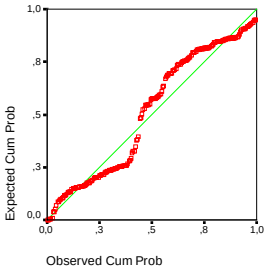
Lognormal P-P Plot of duree interconnexion en heure

LOGIN: gallouin



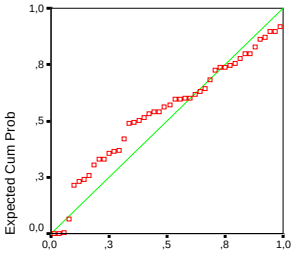
Lognormal P-P Plot of duree interconnexion en

LOGIN: guix



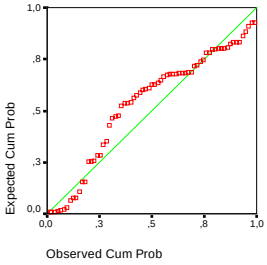
Lognormal P-P Plot of duree interconnexion e

LOGIN: gzing



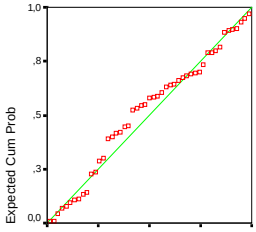
Lognormal P-P Plot of duree interconnexion

LOGIN: hedi.dc



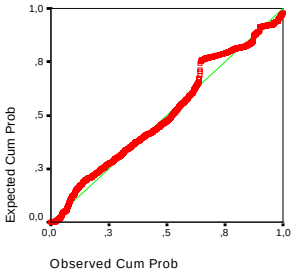
Lognormal P-P Plot of duree interconnexion en heure

LOGIN: HERPIN



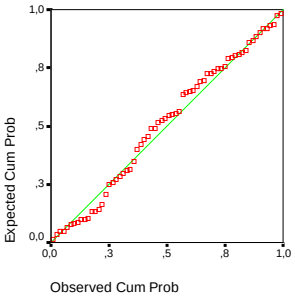
Lognormal P-P Plot of duree inter

LOGIN: jeannot3



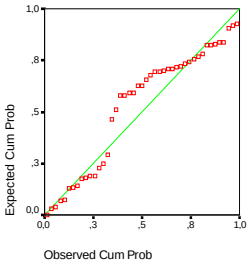
Lognormal P-P Plot of duree interconnexion en he

LOGIN: juboubou



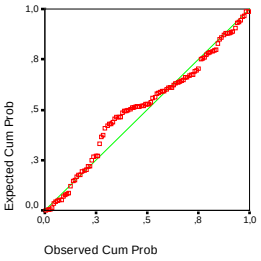
Lognormal P-P Plot of duree interconnexion en heur

LOGIN: juju75



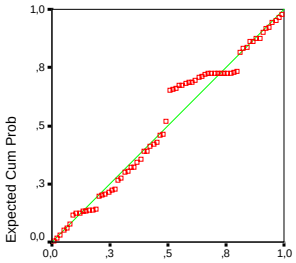
Lognormal P-P Plot of duree interconnexion en heure

LOGIN: kabotine



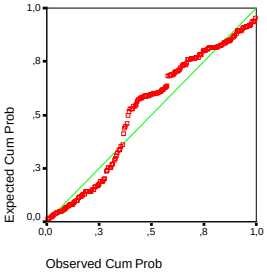
Lognormal P-P Plot of duree interconnexion

LOGIN: Kathrin



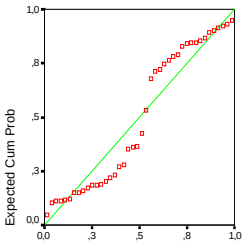
Lognormal P-P Plot of duree interconnexion en h

LOGIN: Kifran



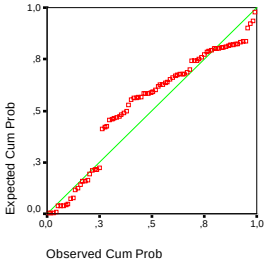
Lognormal P-P Plot of duree interconnexion en heur

LOGIN: levieux



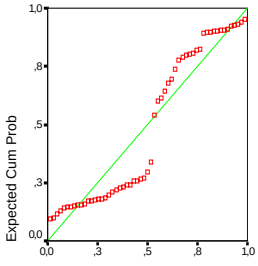
Lognormal P-P Plot of duree interconnexion en heure

LOGIN: Loquet



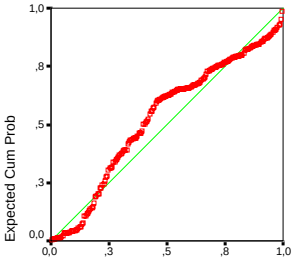
Lognormal P-P Plot of duree interconnexion

LOGIN: m@hy



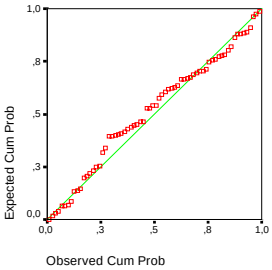
Lognormal P-P Plot of duree interconnexion er

LOGIN: mperrot



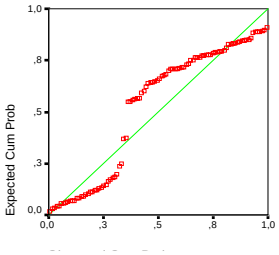
Lognormal P-P Plot of duree interconnexion

LOGIN: Naemesis



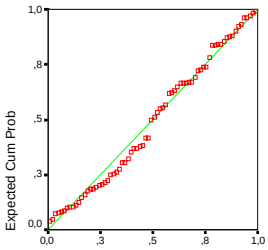
Lognormal P-P Plot of duree interconnexion en heure

LOGIN: nhai07



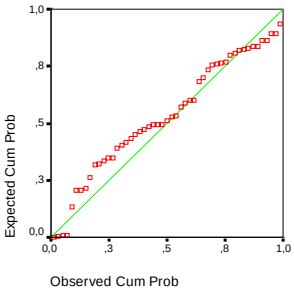
Lognormal P-P Plot of duree interconnexion er

LOGIN: philia



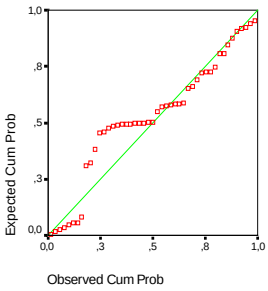
Lognormal P-P Plot of duree interconnexion er

LOGIN: prozaiik



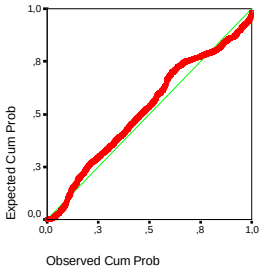
Lognormal P-P Plot of duree interconnexion

LOGIN: RAPH



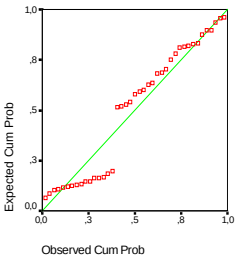
Lognormal P-P Plot of duree interconnexion en heure

LOGIN: semery



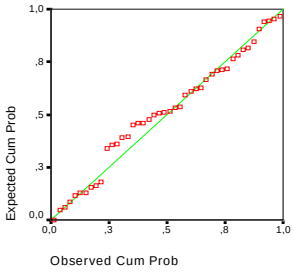
Lognormal P-P Plot of duree interconnexion en

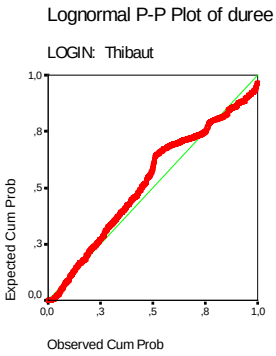
LOGIN: Souris

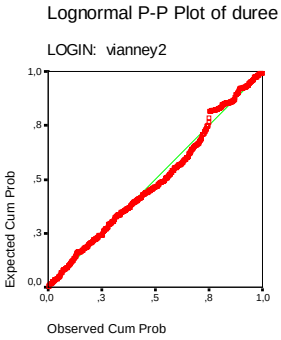


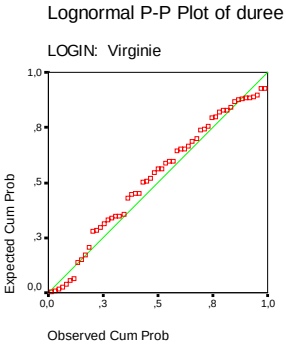
Lognormal P-P Plot of duree interconnexion en h

LOGIN: Stèph





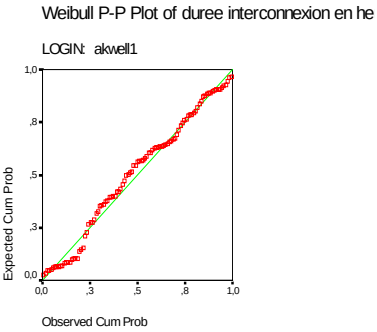


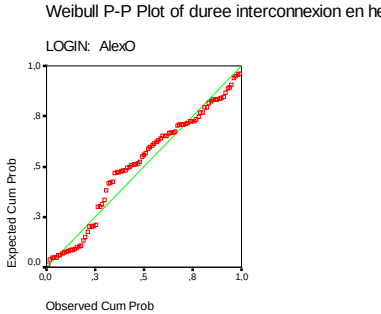


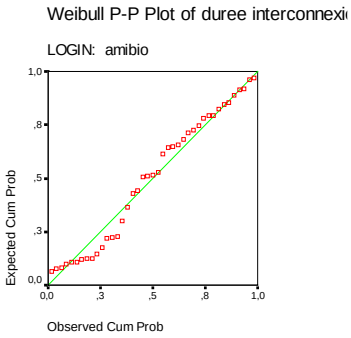
L'ensemble de ces courbes montre que la loi log-normale est mieux représentative des données observées que la loi Gamma (4 cas seulement sur 43 sont particulièrement mal représentés :eb95, levieux, m@hy, nhai07). Cependant, on constate sur plusieurs cas quelques écarts entre le modèle et les données observées (comme AlexO ou bgougay ar exemple) avec une surestimation dans ces cas de la durée interconnexion par le modèle. En revanche, plusieurs cas mal représentés par la loi Gamma sont mieux représentés par la loi log-normale (ex : AnnA, vianney2).

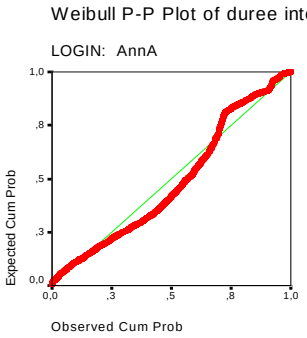
LOI WEIBULL :

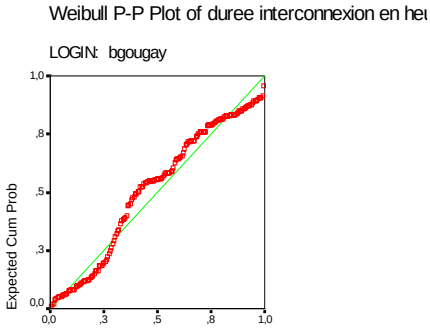
La modélisation par une loi de Weibull donne les résultats suivants (l'ensemble des résultats est fourni en annexe 2):

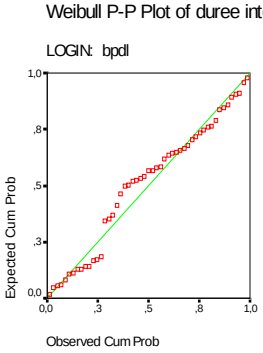


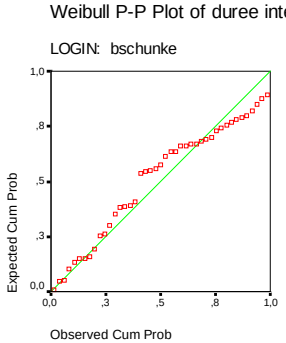


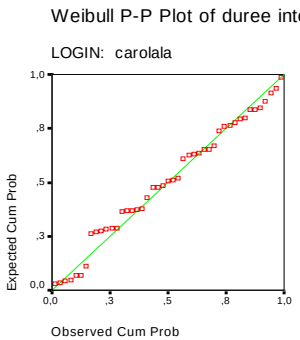






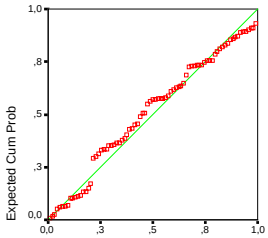






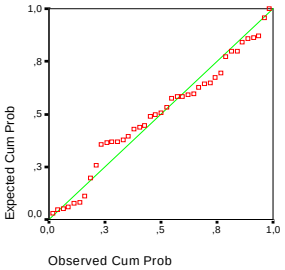
Weibull P-P Plot of duree interconnexion

LOGIN: Delatte



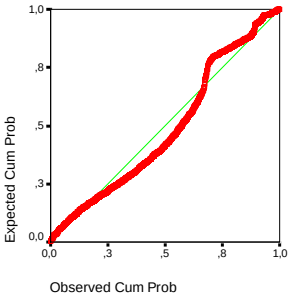
Weibull P-P Plot of duree interconnexion en heure

LOGIN: didine77



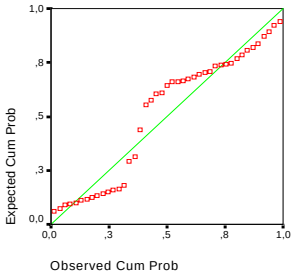
Weibull P-P Plot of duree interconnexion en heure

LOGIN: djoule



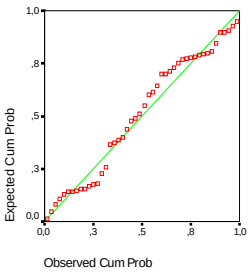
Weibull P-P Plot of duree interconnexion en heure

LOGIN: eb95



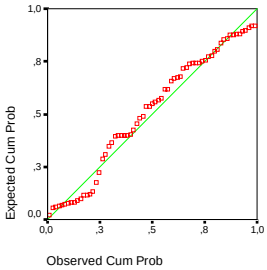
Weibull P-P Plot of duree interconnexion en he

LOGIN: fallouis



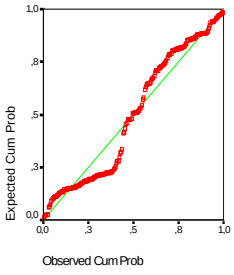
Weibull P-P Plot of duree interconnexion en heure

LOGIN: gallouin



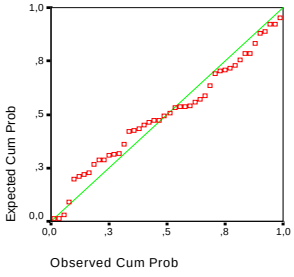
Weibull P-P Plot of duree interconnexion en heur

LOGIN: guix



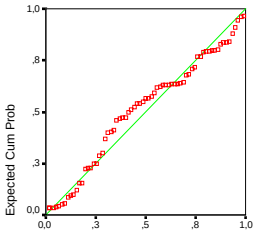
Weibull P-P Plot of duree interconnexion en heure

LOGIN: gzing



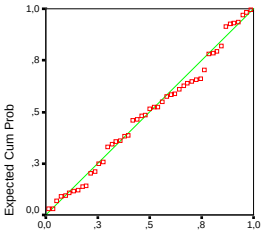
Weibull P-P Plot of duree interconnexion en heure

LOGIN: hedi.dc



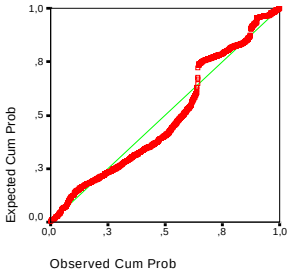
Weibull P-P Plot of duree interconnexion en heure

LOGIN: HERPIN



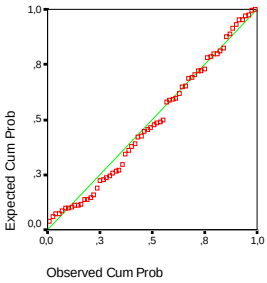
Weibull P-P Plot of duree interconnexion en

LOGIN: jeannot3



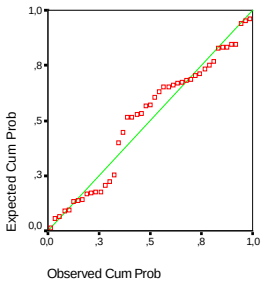
Weibull P-P Plot of duree interconnexion en heure

LOGIN: juboubou



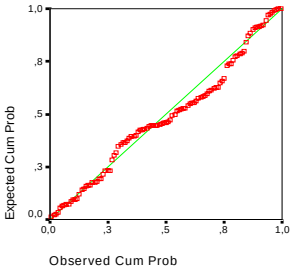
Weibull P-P Plot of duree interconnexion e

LOGIN: juju75



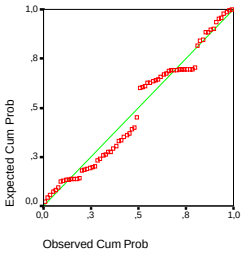
Weibull P-P Plot of duree interconnexion en heure

LOGIN: kabotine



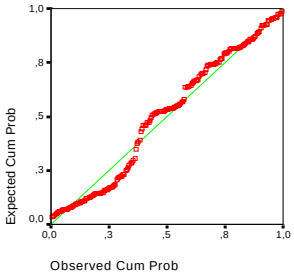
Weibull P-P Plot of duree interconnexion en

LOGIN: Kathrin



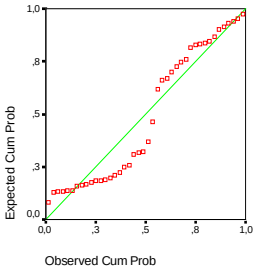
Weibull P-P Plot of duree interconnexion en heure

LOGIN: Kifran



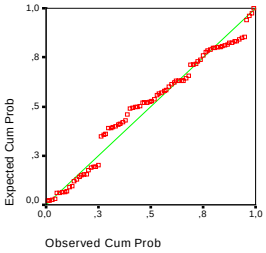
Weibull P-P Plot of duree interconnexion en heu

LOGIN: levieux



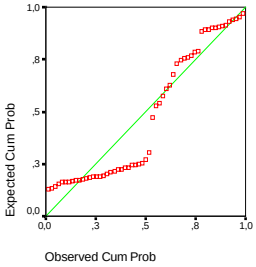
Weibull P-P Plot of duree interconnexion en heure

LOGIN: Loquet



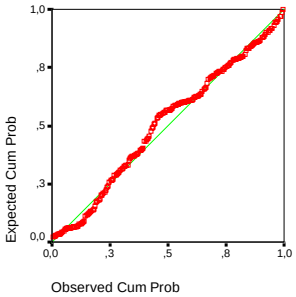
Weibull P-P Plot of duree interconnexion e

LOGIN: m@hy



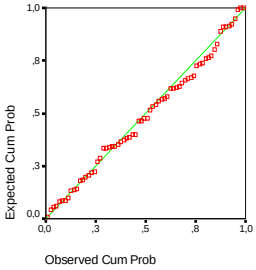
Weibull P-P Plot of duree interconnexion en heure

LOGIN: mperrot



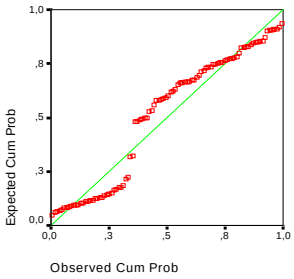
Weibull P-P Plot of duree interconnexion en heure

LOGIN: Naemesis



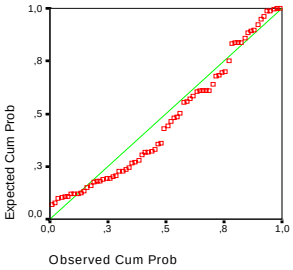
Weibull P-P Plot of duree interconnexion en

LOGIN: nhai07



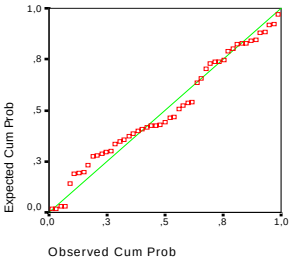
Weibull P-P Plot of duree interconnexion e

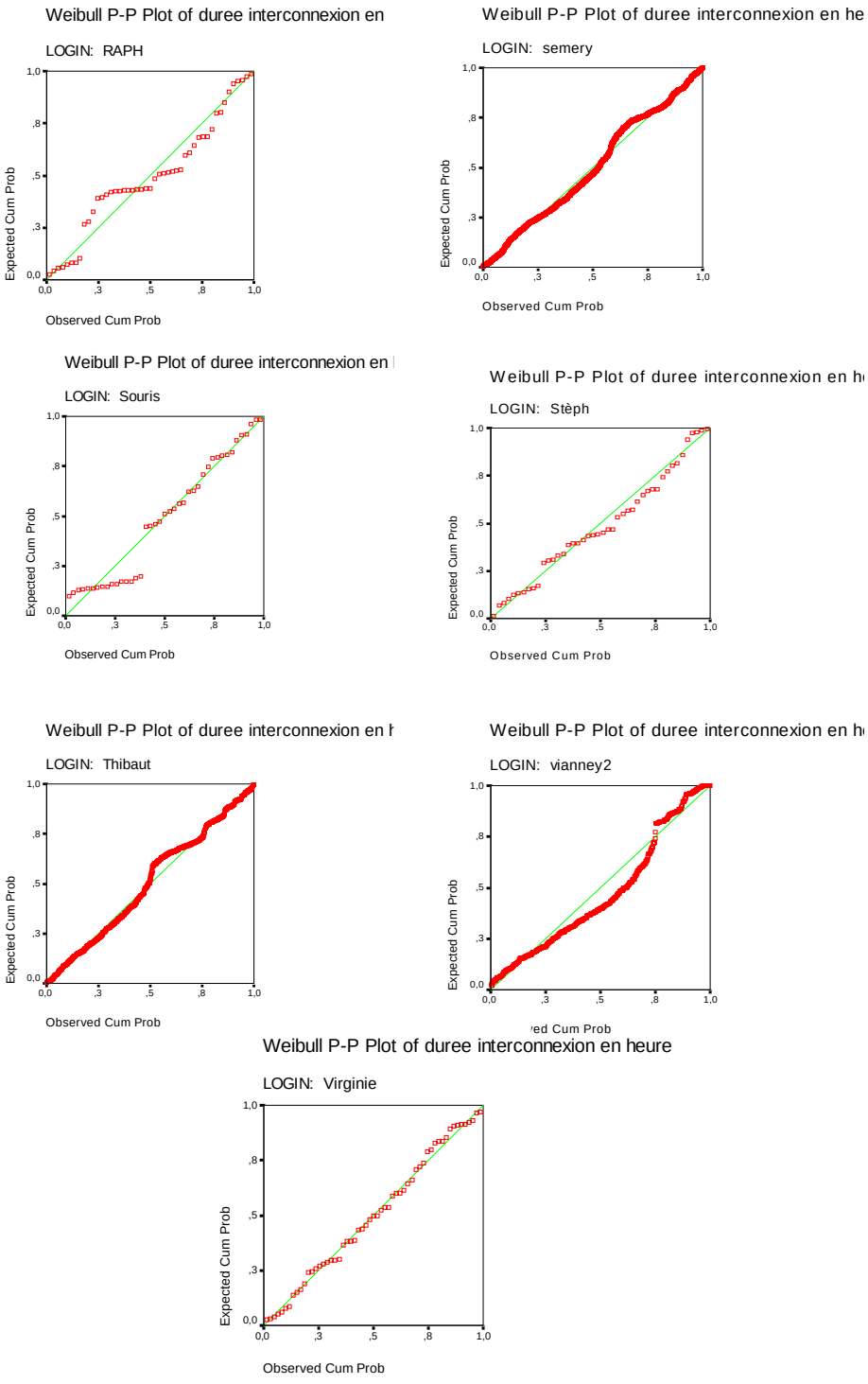
LOGIN: philia



Weibull P-P Plot of duree interconnexion en

LOGIN: prozaiik





L'ensemble de ces graphiques montrent que que la loi de Weibull est plus proche des données réelles que les loi Gamma ou Log-normale. Excepté les cas Souris, m@hy ou levieux, la loi de Weibull représente bien les données observées.

Rappelons ici les formules des trois lois :

Distribution de la loi Gamma:

$$f(X)=\frac{\lambda^{c} X^{c-1} e^{-\lambda X}}{\Gamma(c)}$$

avec λ et c paramètres positifs du modèle

La loi Gamma est une loi flexible qui permet de représenter de nombreux phénomènes. Lorsque c=1, la loi Gamma se réduit à une loi exponentielle.

Distribution de la loi Log-Normale:

$$f(X)=\frac{1}{X\sigma(2\pi)^{1/2}} e^{\left[-[\log(X/m)]^{2}/2\sigma^{2}\right]}$$

m, σ^2 sont les paramètres du modèle

La loi log-normale représente une variable aléatoire dont le logarithme suit une loi normale.

Distribution de la loi Weibull :

Cette distribution est caractérisée par :

$$f(X)=c X^{c-1} \lambda^{c} e^{-(X\lambda)^{c}} \qquad 0 \leq X \leq +\infty$$

avec :

c : paramètre de forme (shape parameter)

λ : paramètre d'échelle (scale parameter)

Notons que la distribution Weibull est une forme généralisée de la distribution exponentielle (quand c=1, Weibull se réduit à une distribution exponentielle ; quand c=2, la distribution de Weibull est souvent connue sous le nom de distribution de Raleigh).

On remarque que la loi Weibull est la loi qui présente la plus simple formule mathématique, ce qui constitue un autre élément justifiant notre préférence pour cette loi.

Cependant, nous ne pouvons statuer entre les trois lois à partir de la simple analyse des courbes PP Plots. Par conséquent, nous avons calculé dans une deuxième étape le maximum de vraisemblance pour chaque loi et pour chacun des 135 cas.

Rappel de la méthode du maximum de vraisemblance :

La fonction de vraisemblance est définie comme la probabilité d'observer tous les points de données. Cette probabilité est calculée à partir du modèle, comme fonction de ses paramètres:

$$L(\text{paramètres}) = p(\text{données}/\text{paramètres})$$

Pour un ensemble de paramètres donné, L (paramètres) donne la probabilité d'obtenir les données réelles. Pour un jeu de données, les paramètres estimés du modèle sont ceux qui maximise cette vraisemblance.

Le maximum de vraisemblance a été calculé pour les 3 lois (Gamma, Log-Normale et Weibull) et pour chacun des 135 agents, sur Excel (via le solveur). La méthodologie suivie est celle recommandée par Hardie et Fader (2002). Les résultats sont fournis dans le tableau ci-dessous.

Tableau 3.3: Comparaison des maximums de vraisemblance par loi

Agent A	Gamma	Weibull	Lognormale
130961	-265,29	-266,1	-274,59
adribou	-31,78	-31,25	-33,98
akwell1	-234,02	-234,42	-240,99
AlexO	-221,67	-222,5	-229,49
ama1	-20,86	-30,54	-9,93
amibio	-78,4	-78,52	-84,68
anakeen	-27,46	-27,64	-34,73
AnnA	-2077,43	-2065,7	-2186,44
apopo	-23,17	-23,44	-27,28
ashley85	-4,37	-4,32	-6,98
Aude	-51,59	-51,83	-54,9
badir	-49,3	-49,32	-51,18
bgougay	-417,3	-418,85	-430,34
blondi	-21,55	-21,61	-24,06
bontze	-11,57	-11,39	-11,1
bpdl	-123,72	-124,46	-128,49
brusson1	-34,87	-34,81	-34,77
bruyere	-50,45	-50,68	-52,03
bschunke	-73,41	-73,28	-73,78
carolala	-94,26	-93,97	-94,77
cathou1	-8,47	-8,48	-8,94
CATHOU25	-14,27	-14,35	-14,11
cclerget	-14,41	-14,48	-14,21
Céline21	-53,64	-53,73	-54,38
chap's	-10,4	-10,39	-12,6
Chrispi1	-10,06	-9,95	-11,5
clemisa	-63,78	-53,14	-54,86
Cypher	-72,97	-66,71	-67,37
DALMATE	-27,98	-27,97	-31,01
Delatte	-184,36	-184,44	-187,46
delillle	-85,85	-85,37	-392,53
didine77	-104,36	-104,42	-105,8
Dinou	-76,73	-77,12	-80,02
djoule	NCP*	-2221,36	-2369,08
dolores3	-25,96	-26,05	-27,44
doppel	-19,23	-19,24	-20,23
dperrot	-5,97	-6,55	-9,86
dzhouari	-13,41	-13,44	-13,76
eb95	-76,18	-76,78	-85,82
eb96	-12,51	-12,55	-13,9
ehran	-64,03	-64,17	-66,14
Emilie	-69,75	-69,55	-73,27
estia	-60,22	-60,34	-65,72
eusebio1	-28,26	-28,16	-29,63
Fab007	-24,94	-24,94	-25,72
fabby93	-13,5	-13,68	-14,53
fabou	-22,69	-22,77	-23,39
fallouis	-79,5	-79,35	-85,3

Agent A	Gamma	Weibull	Lognormale
fdrugeat	-0,01	-0,28	0
fevrard	-38,45	-38,53	-41,35
FOURNIAL	-20,66	-20,65	-21,01
Fruor	-11,74	-10,55	-12,73
frzpower	-16,6	-16,57	-18,37
GaëlleM	-19,2	-19,02	-19,76
gallouin	-141,48	-141,5	-146,48
Ginie31	-21,47	-21,38	-21,33
gougay	-82,89	-83,05	-84,16
Greensky	-34,73	-34,41	-36,04
guix	-353,8	-355,88	-379,26
gzing	-118,54	-118,6	-120,29
hedi.dc	-154,73	-154,38	-156,72
HERPIN	-100,37	-100,01	-104,04
hutch	-83,75	-83,9	-85,29
j-kolb	-8,89	-8,84	-9,32
jeanbask	-23,54	-23,08	-25
jeannot3	-1496,88	-1495,65	-1573,44
jeda	-48,21	-48,2	-49,6
JIMASA	-6,69	-6,54	-6,92
joewj	-48,05	-48,22	-50,63
josetta	-15,97	-15,95	-16,08
juboubou	-144,29	-144,46	-151,22
juju75	-100,81	-101,66	-106,49
kabotine	-252,86	-251,15	-252,09
Kathrin	-134,53	-133,81	-137,41
Kifran	-410,89	-411,66	-428,1
lbarbier	-309,59	-309,23	-320,43
leleu5	-9,61	-9,57	-10,93
levieux	-61,49	-61,68	-75,65
lilimimi	-25,57	-25,17	-25,38
LoïcT	-22,89	-22,94	-23,48
Loquet	-206,44	-206,51	-210,58
luberdov	-3,21	-2,6	-3,93
ludovicb	-12,3	-12,44	-12,99
m@hy	NCP	-94,94	-111,24
manwe	-47,64	-47,8	-49,35
mary_ce	-50,47	-50,32	-57,84
Mayfair	-55,51	-54,81	-61,26
mekiela	-80,93	-80,99	-81,9
Michel	-49,05	-48,97	-49,91
micka74	-28,26	-28,39	-29,04
mike_s	-31,42	-31,49	-32,12
miloumoi	-64,41	-64,18	-64,52
minious	-58,48	-58,54	-65,2
modjo-13	-33,84	-33,63	-34,74
mperrot	-411,62	-411,22	-422,12
Munier	-34,91	-34,79	-38,69

NCP* : le solveur excel ne converge pas

Agent A	Gamma	Weibull	Lognormale
Naemesis	-172,56	-172,62	-175,11
Nathalie	-20,37	-20,3	-20,96
nhai07	-212,6	-213,7	-225,55
nickus	-12,2	-12,19	-12,11
ninie12	-17,7	-17,71	-23,43
noel_net	-15,45	-15,52	-16,39
philia	NCP*	-124,08	-138,02
picardjf	-16,35	-16,35	-16,56
popcorn	-33,63	-33,61	-36,84
poupide	-25,02	-25,09	-25,55
prioton	-37,51	-37,67	-41,77
prozaiik	-120,69	-120,31	-119,68
quentin	-66,28	-66,11	-67,34
RAPH	-112,76	-112,03	-112,23
relindis	-0,44	-0,07	-0,6
rlaval	-3710,84	-1364,33	-1069,24
romain37	-54,78	-54,81	-56,34
sandor	-43,7	-43,21	-46,08
sebulba9	-7,36	-7,45	-7,88
semery	-1823,71	-1813,93	-1874,16
sicart	-9,05	-9,16	-10,36
sophynet	-34,07	-34,08	-37,27
Souris	-81,63	-82,15	-89
starchen	-33,72	-33,86	-75,31
Stèph	-95,57	-95,09	-97,42
stumpf	-60,99	-61,03	-61,84
Sylvie4	-63,01	-62,92	-64,31
Tantan	-8,61	-8,67	-9,82
Thibaut	-1549,74	-1550,14	-1594,86
Tonio2k	-47,58	-47,82	-49,44
toto2	-10,3	-10,35	-11,63
tranb	-63,7	-63,63	-66,32
vianney2	NCP*	-586,46	-591,04
Virginie	-130,95	-130,87	-132,67
xzouf	-5,97	-8,98	-9,86
Yseulys	-1037,3	-1038,82	-1063,89
Yves	-23,64	-23,58	-25,7
zaultar	-4,22	-4,32	-8,97

NCP* : le solveur excel ne converge pas

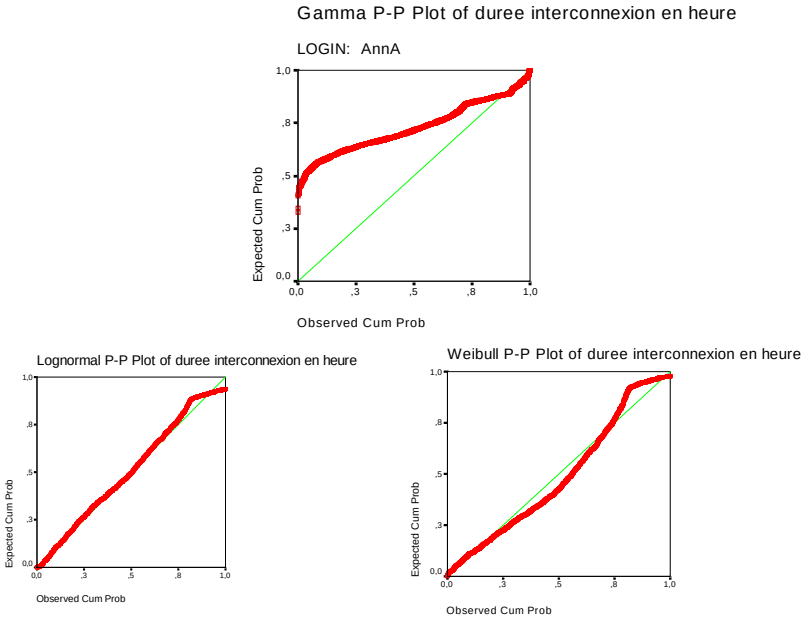
Avant de conclure sur les données de ce tableau, nous nous sommes tout d'abord concentrés sur 12 cas « atypiques » pour lesquels le choix de la loi est différent selon que la sélection s 'effectue à partir du meilleur maximum de vraisemblance ou à partir de la comparaison des courbes PP plots.

Etudions plus précisément ces 12 cas :

Cas 1, Anna :

Les résultats des PP plots sont les suivants :

Figure 3.4: PP plot durée interconnexion login Anna



La comparaison des PP plots amène à choisir la loi Log Normale qui est celle qui représente le mieux les données.

Cependant, les maximums de vraisemblances des 3 lois (issus du tableau 2) sont :

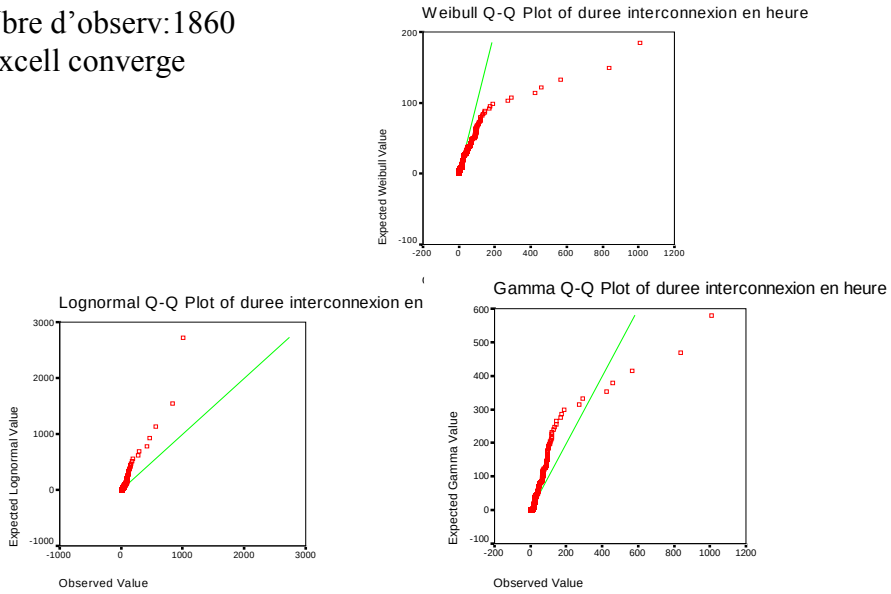
Gamma : -2077,43 Weibull : -2065,7 Lognormale : -2186,44

La loi qui maximise le plus la vraisemblance est la loi de Weibull que nous sélectionnerons en utilisant cette méthode.

Cette contradiction s'explique par le fait que le nombre d'observations sur le login Anna est très élevé (1860 connexions) et que les PP plots sont des courbes cumulées, cachant certains phénomènes. Lorsqu'on représente une comparaison du modèle avec les données observées par un QQ plot (données par quartiles, SPSS divisant les cases en groupes de même nombre de données) des 3 lois pour le login Anna, on obtient :

Figure 3.5: Comparaison des QQ Plots durée interconnexion login Anna

Nbre d'observ:1860
Excell converge



Le QQ plot (par quartiles) versus le PP plot (valeurs cumulées du modèle) est dans ce cas plus précis : on trouve en effet un résultat cohérent avec le maximum de vraisemblance, à savoir que la loi de Weibull est la meilleure des trois lois sur la majorité des données. Ce ne sont que quelques valeurs extrêmes (les 10 points à droite de la courbe, pour lesquels le modèle sous estime les données observées très fortes) qui font dévier la courbe Weibull des données observées (d'où l'écart sur les valeurs cumulées dans le PP plot), alors que la loi Log-Normale donne beaucoup plus de points divergents sur la majorité des données (QQ plot), la courbe en cumulé (PP plot) écrasant ces écarts.

11 autres cas « atypiques » :

Les autres 11 cas atypiques sont les suivants :

Tableau 3.5: Comparaison des maximums de vraisemblance par loi

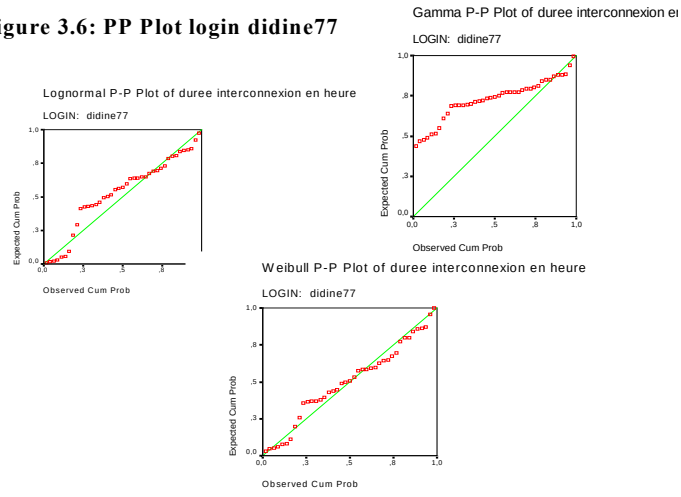
Agent A	Gamma	Weibull	Lognormale	Loi choisie
didine77	-104,36	-104,42	-105,8	Gamma ou Weibull
djoule	NCP*	-2221,36	-2369,08	Weibull
juboubou	-144,29	-144,46	-151,22	Gamma ou Weibull
kabotine	-252,86	-251,15	-252,09	Weibull
Loquet	-206,44	-206,51	-210,58	Gamma ou Weibull
m@hy	NCP*	-94,94	-111,24	Weibull
mperrot	-411,62	-411,22	-422,12	Gamma ou Weibull
Naemesis	-172,56	-172,62	-175,11	Gamma ou Weibull
nhai07	-212,6	-213,7	-225,55	Gamma
philia	NCP*	-124,08	-138,02	Weibull
vianney2	NCP*	NCP*	-591,04	Log Normale

NCP* : le solveur excel ne converge pas

Nous allons détailler 2 de ces cas, les autres étant similaires en terme de résultat.

Le cas *didine77* montre les caractéristiques suivantes :

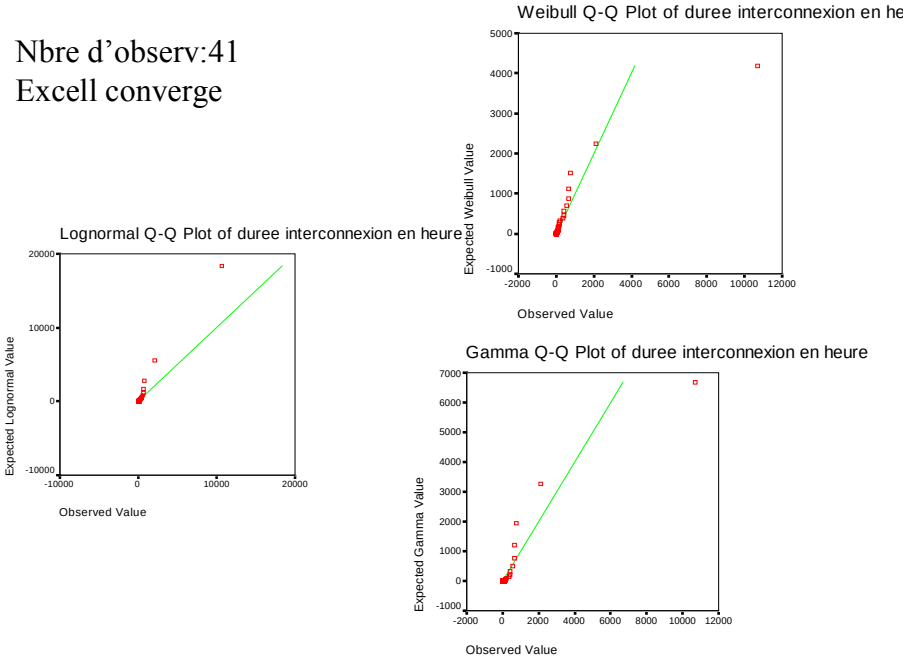
Figure 3.6: PP Plot login didine77



La figure 3.6 présentant les PP plot de la durée interconnexion du login didine77 ferait plutôt choisir la loi de Weibull et rejeter Gamma, résultats en partie contradictoires avec la loi Gamma (ou Weibull) sélectionnée par le maximum de vraisemblance.

Figure 3.7: login didine77 comparaison des QQ Plots

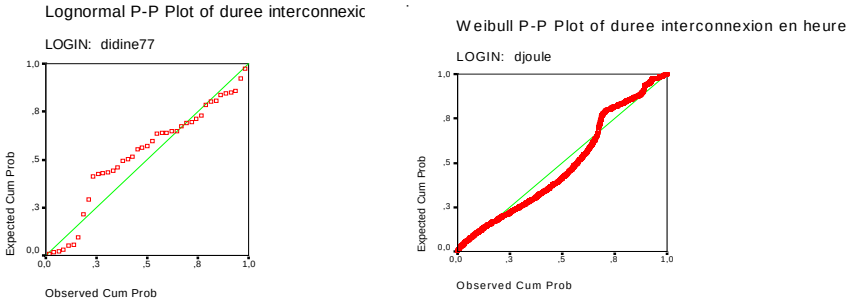
Nbre d'observ:41
Excell converge



Comme pour le login Anna, l'analyse des QQ plot montre que c'est un point extrême (forte durée interconnexion sous estimé pr le modèle) qui en valeurs cumulées fait dévier sur le PP plot la loi de Weibull des données observées. Mais la comparaison des QQ plot entraîne le choix de Weibull et le rejet de Gamma et de la loi Log Normale.

Le cas du login *djoule* est le suivant :

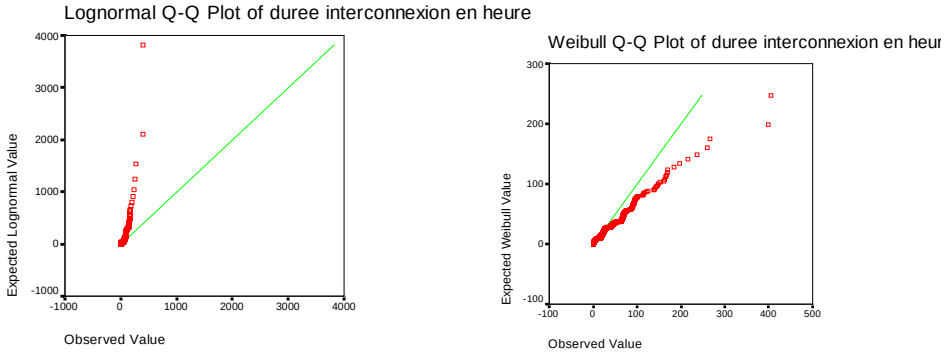
Figure 3.8: PP plot login djoule



Dans le cas du login djoule, le solveur excel ne converge pas pour la loi Gamma et l'analyse du maximum de vraisemblance donne la loi Weibull comme meilleure loi (tableau 3.5), alors que les PP plots montrent des lois Weibull et Log Normale proches.
L'analyse des QQ plots (Figure 10) confirme le choix de Weibull et l'analyse du maximum de vraisemblance en rejetant la loi Log-normale.

Figure 3.9 : Login djoule
Comparaison des QQ Plots

Nbre d'observ: 1831
Excell: Gamma seul
ne converge pas



L'analyse des QQ plots des 9 autres cas, comme les 2 présentés ci-dessus, confirment les résultats de la méthode du maximum de vraisemblance. Le choix final des lois pour les 12 cas atypiques est donc celui du tableau 3.5 précédemment présenté. Le détail de ces 9 autres cas est fourni en annexe 3.

Reprenons maintenant l'analyse de l'ensemble des résultats du tableau 3.3 : les maximums de vraisemblances sont très proches d'une loi à l'autre. Chacune des lois ayant le même nombre de paramètres (2), le choix de la loi s'effectue en sélectionnant la loi qui donne le plus grand maximum de vraisemblance.
Une synthèse de la comparaison des maximums de vraisemblance fournie dans le tableau 3.3 précédent est décrite dans le tableau 3.6 ci-dessous :

Tableau 3.6 : nbre de cas où la loi fournit le plus fort maximum de vraisemblance

	Count	Count	Count
Gamma	66		
Log Normale			11
Weibull		58	

On constate que dans 66 cas, la loi Gamma fournit le plus grand maximum de vraisemblance, tandis que la loi Weibull fournit le plus grand maximum de vraisemblance dans 58 cas, 11 uniquement pour la loi Log-Normale, ce qui devrait nous orienter vers la loi Gamma. Cependant, l'arbitrage entre la loi Gamma et la loi Weibull demande d'autres analyses. Le tableau 3.7 ci-dessous donne la répartition des écarts entre le maximum de vraisemblance fourni par la loi Gamma et la loi Weibull. On constate que dans 90% des cas, l'écart est compris entre –0,56 et +0,56, soit une différence extrêmement faible.

Tableau 3.7 : statistiques sur la différence
Max vraisemblance Gamma-Max vraisemblance Weibull

N	Valid	135
	Missing	0
Mean		-19,1053
Median		,0151
Std. Deviation		203,48743
Minimum		-2346,51
Maximum		32,22
Percentiles	10	-,5676
	20	-,3533
	30	-,1156
	40	-,0347
	50	,0151
	60	,0634
	70	,0930
	80	,1719
	90	,5619

Pour finalement sélectionner le meilleur modèle entre Weibull, Log-Normale et Gamma, nous avons recherché une méthode plus robuste. Nous avons choisi d'appliquer le test de Vuong (Econometrica, 1989), qui n'a à notre connaissance jamais été appliqué à date dans le

domaine du marketing. Ce test permet de choisir le meilleur modèle parmi plusieurs lois non « nested », c'est à dire non intégrées l'une dans l'autre. Ce test correspond parfaitement à notre situation. En effet, la loi Log-Normale n'est pas un cas particulier des lois Weibull ou Gamma. De même la loi Gamma n'est pas un cas particulier des lois Weibull et Log Normale. Enfin, la loi Weibull n'est pas non plus un cas particulier des deux autres lois. On est donc bien dans le cas de la comparaison de 3 modèles non intégrés l'un dans l'autre. Comme nous le verrons dans les chapitres suivants de la thèse, dans de nombreux cas d'application en marketing, on compare des modèles qui sont intégrés l'un dans l'autre en appliquant la méthode du ratio de vraisemblance, méthode qui ne peut pas être appliquée ici pour nos données.

Test de Vuong :

Dans son article (Econometrica, 1989), Vuong développe son test fondé sur le critère KLIC (Kullbach-Leibler (1951) Information Criterion) qui mesure la distance entre une distribution donnée et la distribution réelle des données. Le principe du test de Vuong est que le meilleur modèle sera celui qui minimisera le critère KLIC.

Principe du test de Vuong :
Considérons les 2 modèles conditionnels suivants:

$$F_\theta = \{f(y|z;\theta);\ \theta \in \Theta\} \text{ de densité conditionnelle } h^0(y|z)$$
$$G_\gamma = \{g(y|z;\gamma);\ \gamma \in \Gamma\}$$

La statistique du ratio de maximum de vraisemblance du modèle F_θ par rapport au modèle G_γ est :

$$LR_n(\hat{\theta}_n,\hat{\gamma}_n) \equiv L^f_n(\hat{\theta}_n) - L^g_n(\hat{\gamma}_n) = \sum_{t=1}^{n} \log \frac{f(Yt/Zt;\hat{\theta}_n)}{g(Yt/Zt;\hat{\gamma}_n)}$$

Où $\hat{\theta}_n$ et $\hat{\gamma}_n$ sont les estimateurs des paramètres par le maximum de vraisemblance.

Le test de Vuong est le suivant :

On teste l'hypothèse H_0 qui est l'équivalence des deux modèles, contre H_f qui est l'hypothèse selon laquelle le modèle F_θ est meilleur que le modèle G_γ, et l'hypothèse H_g selon laquelle le modèle G_γ est meilleur que le modèle F_θ .

D'après Vuong, si F_θ et G_γ sont strictement non « nested » (non intégrés), alors :

(i) Sous H_0 : $n^{-1/2}\ LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n \rightarrow N(0, 1)$

(ii) Sous H_f : $n^{-1/2}\ LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n \rightarrow +\infty$

(iii) Sous H_g : $n^{-1/2}\ LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n \rightarrow -\infty$

avec

$$\hat{\omega}_n \equiv \frac{1}{n}\sum_{t=1}^{n}\left[\log\frac{f(Yt/Zt;\hat{\theta}n)}{g(Yt/Zt;\hat{\gamma}n)}\right]^2 - \left[\frac{1}{n}\sum_{t=1}^{n}\log\frac{f(Yt/Zt;\hat{\theta}n)}{g(Yt/Zt;\hat{\gamma}n)}\right]^2$$

Ce théorème fournit des tests simples pour sélectionner le meilleur modèle. On choisit tout d'abord une valeur critique c pour la distribution normale à un certain niveau de signification (par exemple c=1,96 avec un degré de signification de 5%).

Si la valeur de la statistique $n^{-1/2}\ LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n$ est supérieure à c, on rejette l'hypothèse nulle H_0 (les modèles ne sont donc pas équivalents) en faveur de H_f, où le modèle F_θ est meilleur que le modèle G_γ.

Si en revanche $n^{-1/2}\ LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n$ est inférieur à -c, on rejette l'hypothèse nulle H_0 en faveur de H_g, où le modèle G_γ est meilleur que le modèle F_θ.

Si $| n^{-1/2}\ LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n | \leq c$, on ne peut pas différencier les deux modèles.

La statistique $n^{-1/2}\ LR_n(\hat{\theta}_n, \hat{\gamma}_n) / \hat{\omega}_n$ est simple à calculer, car :

$$n^{1/2}\tilde{\omega}_n = \sum m_t^2 \text{ où } \quad m_t \equiv \log[f(Y_t|Z_t;\hat{\theta}_n)/g(Y_t|Z_t;\hat{\gamma}_n)]$$

Le tableau 3.8 ci-dessous détaille donc le calcul de la statistique de Vuong et la comparaison des lois Weibull, Gamma et log-Normale pour chacun des 135 agents. La synthèse des résultats est fournie dans le tableau 3.9.

Tableau 3.8: Statistique de Vuong pour le choix de la meilleure loi

130961	n(-1/2) LR / w =	3,96	G	19,80	G	28,26	W	G
adribou	n(-1/2) LR / w =	-0,61	No diffe	4,06	G	4,16	W	No diff GW
akwell1	n(-1/2) LR / w =	2,46	G	18,55	G	23,10	W	G
AlexO	n(-1/2) LR / w =	5,07	G	20,22	G	24,79	W	G
ama1	n(-1/2) LR / w =	3,81	G	-6,02	LN	-18,11	LN	LN
amibio	n(-1/2) LR / w =	0,70	No diffe	9,02	G	8,27	W	No diff GW
anakeen	n(-1/2) LR / w =	0,86	No diffe	4,14	G	3,70	W	No diff GW
AnnA	n(-1/2) LR / w =	-18,17	W	71,87	G	67,13	W	W
apopo	n(-1/2) LR / w =	1,13	No diffe	3,78	G	3,12	W	No diff GW
ashley85	n(-1/2) LR / w =	-0,08	No diffe	0,70	No diffe	0,61	No diffe	No diff GWLN
Aude	n(-1/2) LR / w =	1,94	No diffe	6,14	G	5,21	W	No diff GW
badir	n(-1/2) LR / w =	0,14	No diffe	4,33	G	4,33	W	No diff GW
bgougay	n(-1/2) LR / w =	8,34	G	28,29	G	39,24	W	G
blondi	n(-1/2) LR / w =	0,35	No diffe	2,71	G	2,44	W	No diff GW
bontze	n(-1/2) LR / w =	-0,84	No diffe	-0,93	No diffe	-0,83	No diffe	No diff GWLN
bpdl	n(-1/2) LR / w =	5,60	G	13,29	G	15,98	W	G
brusson1	n(-1/2) LR / w =	-0,74	No diffe	-0,23	No diffe	-0,11	No diffe	No diff GWLN
bruyere	n(-1/2) LR / w =	2,90	G	5,48	G	5,92	W	G
bschunke	n(-1/2) LR / w =	-1,62	No diffe	1,30	No diffe	2,05	W	W
carolala	n(-1/2) LR / w =	-2,52	W	1,55	No diffe	3,06	W	W
cathou1	n(-1/2) LR / w =	0,02	No diffe	0,66	No diffe	0,65	No diffe	No diff GWLN
CATHOU25	n(-1/2) LR / w =	0,16	No diffe	-0,12	No diffe	-0,13	No diffe	No diff GWLN
cclerget	n(-1/2) LR / w =	0,21	No diffe	-0,19	No diffe	-0,20	No diffe	No diff GWLN
Céline21	n(-1/2) LR / w =	1,14	No diffe	2,06	G	2,31	W	No diff GW
chap's	n(-1/2) LR / w =	-0,09	No diffe	1,19	No diffe	1,18	No diffe	No diff GWLN
Chrispi1	n(-1/2) LR / w =	-0,15	No diffe	1,69	No diffe	1,00	No diffe	No diff GWLN
clemisa	n(-1/2) LR / w =	-3,81	W	-3,16	LN	6,35	W	W
Cypher	n(-1/2) LR / w =	-2,85	W	-2,46	LN	2,86	W	W
DALMATE	n(-1/2) LR / w =	-0,02	No diffe	4,15	G	4,27	W	No diff GW
Delatte	n(-1/2) LR / w =	0,55	No diffe	9,01	G	12,18	W	No diff GW
delillle	n(-1/2) LR / w =	-2,52	W	10,94	G	10,95	W	W
didine77	n(-1/2) LR / w =	0,42	No diffe	3,51	G	4,86	W	No diff GWLN
Dinou	n(-1/2) LR / w =	2,86	G	9,06	G	9,02	W	G
djoule	n(-1/2) LR / w =	-80,65	W	-17,28	LN	100,06	W	W
dolores3	n(-1/2) LR / w =	1,01	No diffe	2,73	G	2,37	W	No diff GW
doppel	n(-1/2) LR / w =	0,40	No diffe	1,50	No diffe	1,43	No diffe	No diff GWLN
dperrot	n(-1/2) LR / w =	0,27	No diffe	0,98	No diffe	1,73	No diffe	No diff GWLN
dzhouari	n(-1/2) LR / w =	0,47	No diffe	1,24	No diffe	1,33	No diffe	No diff GWLN
eb95	n(-1/2) LR / w =	3,34	G	10,22	G	8,46	W	G
eb96	n(-1/2) LR / w =	0,36	No diffe	1,21	No diffe	1,11	No diffe	No diff GWLN
ehran	n(-1/2) LR / w =	1,12	No diffe	5,80	G	6,08	W	No diff GW
Emilie	n(-1/2) LR / w =	-0,89	No diffe	6,75	G	7,29	W	No diff GW
estia	n(-1/2) LR / w =	0,90	No diffe	6,52	G	5,97	W	No diff GW
eusebio1	n(-1/2) LR / w =	-0,51	No diffe	2,53	G	2,70	W	No diff GW
Fab007	n(-1/2) LR / w =	-0,02	No diffe	1,94	No diffe	1,64	No diffe	No diff GWLN
fabby93	n(-1/2) LR / w =	1,20	No diffe	2,39	G	1,97	W	No diff GW
fabou	n(-1/2) LR / w =	0,96	No diffe	2,06	G	1,70	No diffe	No diff GW
fallouis	n(-1/2) LR / w =	-1,12	No diffe	7,69	G	7,53	W	No diff GW

fdrugeat	n(-1/2) LR / w =	-0,21	No diffe	-0,02	No diffe	-0,01	No diffe	No diff GWLN
fevrard	n(-1/2) LR / w =	0,76	No diffe	3,97	G	3,60	W	No diff GW
FOURNIAL	n(-1/2) LR / w =	-0,10	No diffe	1,00	No diffe	1,41	No diffe	No diff GWLN
Fruor	n(-1/2) LR / w =	-0,53	No diffe	0,90	No diffe	1,28	No diffe	No diff GWLN
frzpower	n(-1/2) LR / w =	-0,69	No diffe	1,52	No diffe	1,55	No diffe	No diff GWLN
GaëlleM	n(-1/2) LR / w =	-0,37	No diffe	0,50	No diffe	1,14	No diffe	No diff GWLN
gallouin	n(-1/2) LR / w =	0,10	No diffe	12,01	G	12,62	W	No diff GW
Ginie31	n(-1/2) LR / w =	-1,08	No diffe	-0,48	No diffe	-0,22	No diffe	No diff GWLN
gougay	n(-1/2) LR / w =	2,80	G	3,81	G	3,98	W	G
Greensky	n(-1/2) LR / w =	-0,54	No diffe	1,07	No diffe	2,47	W	W
guix	n(-1/2) LR / w =	7,54	G	39,99	G	50,57	W	G
gzing	n(-1/2) LR / w =	1,66	No diffe	4,40	G	4,28	W	No diff GW
hedi.dc	n(-1/2) LR / w =	-1,84	No diffe	4,39	G	8,06	W	No diff GW
HERPIN	n(-1/2) LR / w =	-1,57	No diffe	7,64	G	9,99	W	No diff GW
hutch	n(-1/2) LR / w =	1,21	No diffe	4,21	G	5,44	W	No diff GW
j-kolb	n(-1/2) LR / w =	-0,06	No diffe	0,36	No diffe	0,98	No diffe	No diff GWLN
jeanbask	n(-1/2) LR / w =	-0,40	No diffe	0,77	No diffe	2,53	W	W
jeannot3	n(-1/2) LR / w =	-3,35	W	77,36	G	79,28	W	W
jeda	n(-1/2) LR / w =	-0,07	No diffe	3,61	G	3,77	W	No diff GW
JIMASA	n(-1/2) LR / w =	-0,16	No diffe	0,14	No diffe	0,51	No diffe	No diff GWLN
joewj	n(-1/2) LR / w =	1,00	No diffe	5,53	G	5,10	W	No diff GW
josetta	n(-1/2) LR / w =	-0,17	No diffe	0,36	No diffe	0,52	No diffe	No diff GWLN
juboubou	n(-1/2) LR / w =	0,77	No diffe	14,70	G	16,86	W	No diff GW
juju75	n(-1/2) LR / w =	5,47	G	14,30	G	13,67	W	G
kabotine	n(-1/2) LR / w =	-8,52	W	-1,58	No diffe	2,99	W	W
Kathrin	n(-1/2) LR / w =	-2,78	W	3,96	G	6,83	W	W
Kifran	n(-1/2) LR / w =	3,50	G	37,78	G	47,78	W	G
lbarbier	n(-1/2) LR / w =	-1,41	No diffe	18,54	G	29,06	W	No diff GW
leleu5	n(-1/2) LR / w =	-0,09	No diffe	1,21	No diffe	0,93	No diffe	No diff GWLN
levieux	n(-1/2) LR / w =	1,36	No diffe	8,63	G	8,25	W	No diff GW
lilimimi	n(-1/2) LR / w =	-1,47	No diffe	-0,36	No diffe	0,56	No diffe	No diff GWLN
LoïcT	n(-1/2) LR / w =	0,44	No diffe	1,86	No diffe	2,10	W	W
Loquet	n(-1/2) LR / w =	0,46	No diffe	10,89	G	15,61	W	No diff GW
luberdov	n(-1/2) LR / w =	-0,16	No diffe	1,41	No diffe	10,23	W	W
ludovicb	n(-1/2) LR / w =	1,27	No diffe	1,99	G	1,63	No diffe	G
m@hy	n(-1/2) LR / w =	1,80	No diffe	14,64	G	13,06	W	No diff GW
manwe	n(-1/2) LR / w =	1,26	No diffe	4,91	G	4,91	W	No diff GW
mary_ce	n(-1/2) LR / w =	-0,85	No diffe	4,70	G	5,37	W	No diff GW
Mayfair	n(-1/2) LR / w =	-0,79	No diffe	2,97	G	6,10	W	No diff GW
mekiela	n(-1/2) LR / w =	1,12	No diffe	2,12	G	2,24	W	No diff GW
Michel	n(-1/2) LR / w =	-0,96	No diffe	2,41	G	2,76	W	No diff GW
micka74	n(-1/2) LR / w =	1,57	No diffe	2,49	G	2,79	W	No diff GW
mike_s	n(-1/2) LR / w =	0,66	No diffe	1,63	No diffe	1,96	No diffe	No diff GWLN
miloumoi	n(-1/2) LR / w =	-1,39	No diffe	0,24	No diffe	1,12	No diffe	No diff GWLN
minious	n(-1/2) LR / w =	0,43	No diffe	6,93	G	6,40	W	No diff GW
modjo-13	n(-1/2) LR / w =	-0,91	No diffe	1,89	No diffe	2,98	W	W
mperrot	n(-1/2) LR / w =	-1,76	No diffe	18,67	G	29,93	W	No diff GW
Munier	n(-1/2) LR / w =	-0,91	No diffe	3,70	G	3,69	W	No diff GW
Naemesis	n(-1/2) LR / w =	0,42	No diffe	5,89	G	8,54	W	No diff GW
Nathalie	n(-1/2) LR / w =	1,02	No diffe	1,16	No diffe	8,54	W	W

nhai07	n(-1/2) LR / w =	6,11	G	27,12	G	25,13	W	G
nickus	n(-1/2) LR / w =	-0,10	No diffe	-0,23	No diffe	-0,23	No diffe	No diff GWLN
ninie12	n(-1/2) LR / w =	0,25	No diffe	2,03	G	1,99	W	No diff GW
noel_net	n(-1/2) LR / w =	0,64	No diffe	1,81	No diffe	1,55	No diffe	No diff GWLN
philia	n(-1/2) LR / w =	-1,24	No diffe	17,72	G	18,35	W	No diff GW
picardjf	n(-1/2) LR / w =	-0,10	No diffe	0,65	No diffe	0,57	No diffe	No diff GWLN
popcorn	n(-1/2) LR / w =	-0,16	No diffe	3,78	G	3,58	W	No diff GW
poupide	n(-1/2) LR / w =	0,71	No diffe	1,49	No diffe	1,78	No diffe	No diff GWLN
prioton	n(-1/2) LR / w =	0,97	No diffe	4,70	G	4,13	W	No diff GW
prozaiik	n(-1/2) LR / w =	-3,69	W	-2,72	LN	-2,22	LN	LN
quentin	n(-1/2) LR / w =	-1,47	No diffe	2,93	G	3,69	W	No diff GW
RAPH	n(-1/2) LR / w =	-4,00	W	-1,21	No diffe	0,69	No diffe	W
relindis	n(-1/2) LR / w =	-0,11	No diffe	0,00	No diffe	0,02	No diffe	No diff GWLN
rlaval	n(-1/2) LR / w =	-1054,06	W	#####	LN	-54,55	LN	LN
romain37	n(-1/2) LR / w =	0,24	No diffe	4,55	G	3,67	W	No diff GW
sandor	n(-1/2) LR / w =	-2,56	W	3,47	G	4,11	W	W
sebulba9	n(-1/2) LR / w =	0,25	No diffe	0,37	No diffe	0,40	No diffe	No diff GWLN
semery	n(-1/2) LR / w =	-29,85	W	58,14	G	75,25	W	W
sicart	n(-1/2) LR / w =	0,48	No diffe	1,34	No diffe	1,05	No diffe	No diff GWLN
sophynet	n(-1/2) LR / w =	0,00	No diffe	3,83	G	3,59	W	No diff GW
Souris	n(-1/2) LR / w =	2,82	G	10,94	G	9,34	W	G
starchen	n(-1/2) LR / w =	0,59	No diffe	4,14	G	4,13	W	No diff GW
Stèph	n(-1/2) LR / w =	-2,23	W	4,21	G	7,01	W	W
stumpf	n(-1/2) LR / w =	0,51	No diffe	2,15	G	2,56	W	No diff GW
Sylvie4	n(-1/2) LR / w =	-0,83	No diffe	3,61	G	4,22	W	No diff GW
Tantan	n(-1/2) LR / w =	0,32	No diffe	1,12	No diffe	0,96	No diffe	No diff GWLN
Thibaut	n(-1/2) LR / w =	1,79	No diffe	74,45	G	101,21	W	No diff GW
Tonio2k	n(-1/2) LR / w =	1,47	No diffe	4,84	G	5,58	W	No diff GW
toto2	n(-1/2) LR / w =	0,36	No diffe	1,18	No diffe	1,03	No diffe	No diff GWLN
tranb	n(-1/2) LR / w =	-0,62	No diffe	5,04	G	5,07	W	No diff GW
vianney2	n(-1/2) LR / w =	13,31	G	50,57	G	1,96	No diffe	G
Virginie	n(-1/2) LR / w =	-0,61	No diffe	4,98	G	7,22	W	No diff GW
xzouf	n(-1/2) LR / w =	0,27	No diffe	0,98	No diffe	1,70	No diffe	No diff GWLN
Yseulys	n(-1/2) LR / w =	7,97	G	50,61	G	72,57	W	G
Yves	n(-1/2) LR / w =	-0,90	No diffe	2,01	G	2,06	W	No diff GW
zaultar	n(-1/2) LR / w =	0,00	No diffe	0,12	No diffe	0,32	No diffe	No diff GWLN

Tableau 3.9: Choix final de la meilleure loi

		Frequency	Percent	Valid	Cumulative Percent
Valid	G	17	12,7	12,7	12,7
	LN	3	2,2	2,2	14,9
	No diff GW	57	41,8	41,8	56,7
	No diff GWLN	38	28,4	28,4	85,1
	W	20	14,9	14,9	100,0
	Total	135	100,0	100,0	

Le tableau 3.9 résume les résultats obtenus. On constate que :

- La loi log-normale n'est pas à conserver : seulement 2,2% des cas montrent qu'elle est la meilleure des trois lois
- La loi Weibull est la meilleure dans 14,9% des cas, la loi Gamma dans 12,7% des cas
- Dans 28,4% des cas, il n'y a pas de différence entre les trois lois (No diff GWLN dans le tableau)
- Dans 41,8% des cas, les lois de Weibull et de Gamma sont meilleures que la loi Log Normale, mais elles sont équivalentes entre elles (no diff GW dans le tableau)

En conclusion, on peut donc choisir, pour modéliser la durée interconnexion des agents, soit la loi Gamma, soit la loi Weibull. Etant donné que la loi Weibull est plus simple mathématiquement que la loi Gamma, nous sélectionnons au final cette loi.
Nous avons par conséquent ci-dessus montré que la durée interconnexion des agents A peut être modélisée par une distribution de Weibull, caractérisée par :

$$f(X)=c X^{c-1} \lambda^{c} e^{-(X\lambda)^{c}} \qquad 0 \leq X \leq +\infty$$

avec :
c : paramètre de forme (shape parameter)
λ : paramètre d'échelle (scale parameter)

Notre résultat est innovateur par rapport aux résultats issus de la littérature en Marketing : dans la majorité des articles, les durées inter-visites sur un site Web sont modélisées par une loi exponentielle (Moe and Fader, 2004a) que nous rejetons. Nous modélisons ces durées inter-visites par une loi de Weibull, qui est plus appropriée et sélectionnée grâce au test de Vuong, à notre connaissance pas utilisé jusque là dans la littérature en Marketing.

3.4.2 Loi de durée interconnexion des agents B :

L'objectif de cette partie est d'analyser le comportement de connexion des agents B (amis ou connaissances de l'agent A) lorsqu'ils vont consulter la wish list de l'agent A qu'ils connaissent. Ceci revient à modéliser la loi du franchissement de la transition *connexionB* du réseau de Petri de la figure 3.2, c'est à dire la loi de la variable durée interconnexion des

agents B sur le site www.Millemercis.com. Plus précisément, étant donné que nous modélisons via le réseau de Petri décrit dans la figure 3.2 le comportement sur les listes de chaque agent A (ou login), nous étudions ici la loi de la durée entre deux connexions d'un agent B sur l'une des listes d'un même agent A.

Nous avons appliqué dans cette partie la même méthodologie sur dans la partie 3.4.1 précédente.

Dans un premier temps, nous avons regardé les PP plots de différentes lois, comparant la performances fournies par ces lois pour représenter la durée interconnexion de B (pour un login de l'agent A)aux données observées. Comme précédemment, nous avons rejeté la loi exponentielle dont les résultats ne reflétaient pas les données observées, et avons conservé les lois Gamma, Weibull et Log-normale.

L'ensemble des PP plots pour la durée interconnexion de l'agent B et pour chacune de ces trois lois est fournie dans l'annexe 4. Comme le montre l'annexe 4, pour chaque login, l'une des trois lois est proche des données observées. Conformément à la méthodologie suivie dans la partie 3.3.2, nous avons ensuite effectué un test de Vuong afin de déterminer, sur l'ensemble des logins, la meilleure des lois entre Gamma, Log-normale et Weibull.

Les résultats de ce test sont fournis dans le tableau 3.10 ci-dessous :

Tableau 3.10

Statistique de Vuong pout le choix de la meilleure loi

		G vs W		G vs LN		W vs LN		
130961	n(-1/2) LR / w =	-0,0665	No diffe	28,9365	G	35,58562	W	No diff GW
adribou	n(-1/2) LR / w =	0,22069	No diffe	3,47068	G	5,097379	W	No diff GW
akwell1	n(-1/2) LR / w =	-2,208	W	14,8538	G	15,39116	W	W
AlexO	n(-1/2) LR / w =	-8,6478	W	-102,77	LN	-91,9106	LN	LN
ama1	n(-1/2) LR / w =	-0,0794	No diffe	-1,96	LN	-2,12022	LN	LN
amibio	n(-1/2) LR / w =	-1,2632	No diffe	-0,5452	No diffe	1,376771	No diffe	No diff GWLN
AnnA	n(-1/2) LR / w =	7,57072	G	55,6443	G	55,68701	W	G
apopo	n(-1/2) LR / w =	0,54951	No diffe	1,40772	No diffe	1,691528	No diffe	No diff GWLN
ashley85	n(-1/2) LR / w =	3,70038	G	10,588	G	13,94386	W	G
Aude	n(-1/2) LR / w =	0,09794	No diffe	23,8063	G	26,12492	W	No diff GW
badir	n(-1/2) LR / w =	-1,8109	No diffe	4,27707	G	4,820483	W	No diff GW
bgougay	n(-1/2) LR / w =	0,25013	No diffe	0,91282	No diffe	0,219585	W	No diff GWLN
blondi	n(-1/2) LR / w =	-1,3364	No diffe	0,52786	No diffe	1,180372	No diffe	No diff GWLN
bontze	n(-1/2) LR / w =	-0,538	No diffe	12,0467	G	6,933871	W	No diff GW
bpdl	n(-1/2) LR / w =	-0,667	No diffe	12,1984	G	6,920842	W	No diff GW
brusson1	n(-1/2) LR / w =	-4,7853	W	4,77087	G	5,221924	W	W
bruyere	n(-1/2) LR / w =	-1,7481	No diffe	-2,5899	LN	-2,6041	LN	LN
carolala	n(-1/2) LR / w =	4,43427	G	6,45805	G	7,246585	W	G
cathou1	n(-1/2) LR / w =	0,09839	No diffe	3,53447	G	3,377729	W	No diff GW
CATHOU25	n(-1/2) LR / w =	1,16053	No diffe	1,45476	No diffe	1,531728	No diffe	No diff GWLN
Céline21	n(-1/2) LR / w =	-3,1021	W	2,07414	G	5,178163	W	W
chap's	n(-1/2) LR / w =	0,01747	No diffe	-0,5476	No diffe	-1,29201	No diffe	No diff GWLN
Chrispi1	n(-1/2) LR / w =	-0,1147	No diffe	0,42255	No diffe	0,413355	No diffe	No diff GWLN
clemisa	n(-1/2) LR / w =	-1,0229	No diffe	2,07725	G	5,627198	W	No diff GW
Cypher	n(-1/2) LR / w =	-0,7277	No diffe	-0,5608	No diffe	1,115132	No diffe	No diff GWLN
DALMATE	n(-1/2) LR / w =	-1,0784	No diffe	5,82446	G	6,263227	W	No diff GW
Delatte	n(-1/2) LR / w =	2,02307	G	16,09	G	14,41724	W	G
didine77	n(-1/2) LR / w =	0,9893	No diffe	2,64237	G	3,765204	W	No diff GW
Dinou	n(-1/2) LR / w =	1,81683	No diffe	11,7525	G	12,36585	W	No diff GW
djoule	n(-1/2) LR / w =	23,5511	G	108,728	G	119,3354	W	G
dolores3	n(-1/2) LR / w =	-1E-07	No diffe	0,03569	No diffe	0,039577	No diffe	No diff GWLN
dperrot	n(-1/2) LR / w =	4,16025	G	8,79704	G	10,52019	W	G
dzhouari	n(-1/2) LR / w =	-0,1081	No diffe	1,29381	No diffe	0,551859	No diffe	No diff GWLN
ehran	n(-1/2) LR / w =	4,17421	G	11,4508	G	9,012575	W	G
Emilie	n(-1/2) LR / w =	-19,323	W	3,65447	G	13,82576	W	W
eusebio1	n(-1/2) LR / w =	0,82442	No diffe	4,60204	G	4,13491	W	No diff GW
fabby93	n(-1/2) LR / w =	-0,0112	No diffe	-1,9989	LN	-1,93417	No diffe	LN
fabou	n(-1/2) LR / w =	0,59749	No diffe	3,34147	G	2,923452	W	No diff GW
fallouis	n(-1/2) LR / w =	1,83387	No diffe	2,80396	G	3,104678	W	No diff GW
fevrard	n(-1/2) LR / w =	0,03593	No diffe	1,30613	No diffe	1,366567	No diffe	No diff GWLN
FOURNIAL	n(-1/2) LR / w =	-0,5078	No diffe	6,0475	G	8,565034	W	No diff GW
Fruor	n(-1/2) LR / w =	0,34366	No diffe	1,38946	No diffe	1,258196	No diffe	No diff GWLN
frzpower	n(-1/2) LR / w =	-1,7735	No diffe	-1,6107	No diffe	-1,19908	No diffe	No diff GWLN
GaëlleM	n(-1/2) LR / w =	0,29964	No diffe	1,76374	No diffe	1,666557	No diffe	No diff GWLN
gallouin	n(-1/2) LR / w =	-6,9544	W	151,051	G	120,4561	W	W
Ginie31	n(-1/2) LR / w =	-3,5834	W	3,93164	G	4,231146	W	W
gougay	n(-1/2) LR / w =	-0,5132	No diffe	-0,4907	No diffe	-0,45801	No diffe	No diff GWLN
Greensky	n(-1/2) LR / w =	0,82673	No diffe	10,5753	G	10,33107	W	No diff GW
gzing	n(-1/2) LR / w =	0,79674	No diffe	16,4398	G	16,04305	W	No diff GW
hedi.dc	n(-1/2) LR / w =	0,62803	No diffe	2,55587	G	2,464524	W	No diff GW
hutch	n(-1/2) LR / w =	-0,8489	No diffe	4,94533	G	4,999503	W	No diff GW
jeanbask	n(-1/2) LR / w =	0,11772	No diffe	1,18459	No diffe	1,002511	No diffe	No diff GWLN
jeannot3	n(-1/2) LR / w =	22,3143	G	44,8471	G	56,52127	W	G
joewj	n(-1/2) LR / w =	-0,4353	No diffe	1,68876	No diffe	2,150543	W	W
juboubou	n(-1/2) LR / w =	0,05196	No diffe	-0,0316	No diffe	-0,03829	No diffe	No diff GWLN
juju75	n(-1/2) LR / w =	-0,8489	No diffe	4,94533	G	4,999503	W	No diff GW
kabotine	n(-1/2) LR / w =	5,30153	W G	49,8745	W	60,85834	W	G
Kathrin	n(-1/2) LR / w =	-1,5723	No diffe	1,39473	No diffe	1,784269	No diffe	No diff GWLN
Kifran	n(-1/2) LR / w =	6,39618	G	29,6229	G	25,18842	W	G
lbarbier	n(-1/2) LR / w =	3,46246	G	21,9885	G	28,71183	W	G
leleu5	n(-1/2) LR / w =	0,88579	No diffe	1,2892	No diffe	1,008507	No diffe	No diff GWLN
levieux	n(-1/2) LR / w =	-0,2317	No diffe	13,2826	G	13,41125	W	No diff GW
LoïcT	n(-1/2) LR / w =	-0,9225	No diffe	4,24649	G	4,635647	W	No diff GW
Loquet	n(-1/2) LR / w =	-0,0737	No diffe	6,7025	G	6,598456	W	No diff GW

luberdov	n(-1/2) LR / w =	-0,1837	No diffe	0,36256	No diffe	1,070601	No diffe	No diff GWLN
ludovicb	n(-1/2) LR / w =	-2,2758	W	4,57024	G	4,826478	W	W
m@hy	n(-1/2) LR / w =	1,18308	No diffe	18,7493	G	20,76905	W	No diff GW
manwe	n(-1/2) LR / w =	0,04635	No diffe	-0,031	G	-0,0355	W	No diff GWLN
mary_ce	n(-1/2) LR / w =	-0,3272	No diffe	-0,2688	G	-0,07946	W	No diff GWLN
Mayfair	n(-1/2) LR / w =	2,52063	G	6,72774	G	5,601056	W	G
mekiela	n(-1/2) LR / w =	-0,6888	No diffe	1,35109	No diffe	2,328468	W	W
Michel	n(-1/2) LR / w =	0,28179	No diffe	3,84293	G	3,662297	W	No diff GW
micka74	n(-1/2) LR / w =	-0,5098	No diffe	2,96541	G	2,957743	W	No diff GW
mike_s	n(-1/2) LR / w =	0,22259	No diffe	6,28948	G	7,151223	W	No diff GW
minious	n(-1/2) LR / w =	0,90212	No diffe	13,8504	G	17,07574	W	No diff GW
mperrot	n(-1/2) LR / w =	0,67044	No diffe	32,9148	G	40,64172	W	No diff GW
Munier	n(-1/2) LR / w =	-2,2476	No diffe	23,9947	G	27,21213	W	W
Naemesis	n(-1/2) LR / w =	1,55446	No diffe	8,55162	G	7,536026	W	No diff GW
nhai07	n(-1/2) LR / w =	-0,6743	No diffe	1,28529	No diffe	1,150704	No diffe	No diff GWLN
nickus	n(-1/2) LR / w =	-0,0169	No diffe	6,8396	G	8,706475	W	No diff GW
noel_net	n(-1/2) LR / w =	1,66752	No diffe	8,17398	G	7,159998	W	No diff GW
philia	n(-1/2) LR / w =	-6,1857	W	15,4937	G	12,73862	W	W
popcorn	n(-1/2) LR / w =	0,75946	No diffe	12,8489	G	14,59074	W	No diff GW
prioton	n(-1/2) LR / w =	0,58331	No diffe	3,52581	G	3,300939	W	No diff GW
prozaiik	n(-1/2) LR / w =	-1,4192	No diffe	41,0762	G	53,56897	W	No diff GW
quentin	n(-1/2) LR / w =	-5,5343	W	12,8903	G	16,28619	W	W
RAPH	n(-1/2) LR / w =	-0,3561	No diffe	2,4293	G	0,748192	No diffe	No diff GW
relindis	n(-1/2) LR / w =	0,38102	No diffe	2,86075	G	W	No diffe	No diff GW
rlaval	n(-1/2) LR / w =	7,24432	G	28,567	G	25,21083	W	G
sandor	n(-1/2) LR / w =	0,28868	No diffe	0,32798	No diffe	0,141287	No diffe	No diff GWLN
sebulba9	n(-1/2) LR / w =	-0,947	No diffe	3,90307	G	2,737533	W	No diff GW
semery	n(-1/2) LR / w =	-2,4653	W	74,4282	G	92,60804	W	W
sicart	n(-1/2) LR / w =	1,01529	No diffe	2,8416	G	3,785394	W	No diff GW
sophynet	n(-1/2) LR / w =	-0,0343	No diffe	3,37406	G	3,163224	W	No diff GW
Stèph	n(-1/2) LR / w =	1,37576	No diffe	11,5783	G	10,6273	W	No diff GW
stumpf	n(-1/2) LR / w =	0,02684	No diffe	0,00429	No diffe	2,31E-05	No diffe	No diff GWLN
Sylvie4	n(-1/2) LR / w =	0,00272	No diffe	0,03236	No diffe	0,342027	No diffe	No diff GWLN
Tantan	n(-1/2) LR / w =		No diffe		No diffe		No diffe	No diff GWLN
Thibaut	n(-1/2) LR / w =	18,1294	G	70,5869	G	93,94757	W	G
Tonio2k	n(-1/2) LR / w =	-1,2139	No diffe	5,37172	G	9,029756	W	No diff GW
toto2	n(-1/2) LR / w =	0,05866	No diffe	4,84079	G	4,605942	W	No diff GW
tranb	n(-1/2) LR / w =	4,42021	G	13,1981	G	11,48845	W	G
vianney2	n(-1/2) LR / w =	0,26822	No diffe	18,8287	G	19,40778	W	No diff GW
Virginie	n(-1/2) LR / w =	7,31461	G	17,8707	G	14,07779	W	G
xzouf	n(-1/2) LR / w =	5,23342	G	49,3363	G	56,86212	W	G
Yseulys	n(-1/2) LR / w =	-5,2992	W	68,3535	G	92,5115	W	W
Yves	n(-1/2) LR / w =	-1,26	No diffe	3,8249	G	4,460227	W	No diff GW

Une synthèse des résultats du test de Vuong est fournie dans le tableau 3.11 ci-dessous :

Tableau 3.11: Synthèse des résultats du test de Vuong

	Frequency	Percent	Valid	Cumulative Percent
G	17	15,	15,	15,
LN	4	3,7	3,7	19,
No diff	45	42,	42,	61,
No diff	27	25,	25,	86,
W	14	13,	13,	100,
Total	107	100,	100,	

On constate que dans 25,2% des cas, il n'y a pas de différence entre les trois lois Gamma, Log-normale et Weibull pour représenter la durée interconnexion des agents B. Dans 42,1% des cas, la loi Gamma et Weibull sont équivalentes, et dans 15,9 % des cas, la loi Gamma est la meilleure des trois lois. La loi de Weibull étant de plus mathématiquement plus simple que la loi Gamma , nous avons retenu la loi de Weibull pour représenter la durée interconnexion des agents B, ce qui est également cohérent avec la loi de Weibull déjà retenu pour modéliser la durée entre deux connexions d'un agent A.

Après avoir modélisé les lois de connexions des agents A et B, nous allons maintenant analyser le comportement d'achat de ces agents afin de déterminer la loi de la transition *purchase* du réseau de Petri de la figure 3.2.

3.4.3 Loi d'achat des agents A ou B

Dans cette partie notre objectif est de modéliser la probabilité d'achat des agents A ou B après chacune de leur connexion sur le site Millemercis.com. Les données portent donc sur 22 356 connexions effectuées sur les listes des 135 agents A. Les connexions peuvent être effectuées par les agents A ou les agents B, puisque les deux types d'agents sont susceptibles d'acheter des produits de la liste.

Une première analyse des données nous a permis d'étudier tout d'abord la distribution du nombre d'items achetés sur l'agent A, l'agent B et au total des deux, comme le montre le tableau 3.12.

Tableau 3.12 : Distribution du nombre d'items achetés après une connexion

Nbre items achetes après la connexion

Count

		Type d'agents		
		A	B	Total
Nbre items achetes après connexion	0	12570	9331	21901
	1	210	111	321
	2	60	36	96
	3	28	14	42
	4	14	5	19
	5	8	4	12
	6	15		15
	7	9	1	10
	8	2		2
	9	6	1	7
	10	4	2	6
	11	1	1	2
	12	1	2	3
	13	4		4
	14	2	2	4
	15	1		1
	16	1		1
	17		1	1
	20		1	1
	21	2		2
	22	1	1	2
	24	1		1
	25	1		1
	26	1		1
	64	1		1
Total		12943	9513	22456

On constate tout d'abord sur le tableau 3.12 de forts atypismes : 0,1% des connexions sont suivies de plus de 10 items achetés après la connexion (le chiffre extrême étant 64 items achetés après une connexion). Ceci pourrait correspondre à un comportement de vidage de la liste ou d'achats simultanés d'items par l'agent A lorsque, au bout d'un certain temps, aucun agent B ne lui a offert de cadeaux de sa liste (ce dernier cas est intéressant d'un point de vue marketing puisqu'alors la liste de cadeaux MilleMercis sert tout simplement de liste d'achats, hypothèses tout à fait plausible selon les équipes de Millemercis).

On note également à partir du tableau 3.12 que les achats au delà de 10 items représentent 459 items achetés soit 30% des achats totaux, ce qui signifie que le comportement qui consiste à acheter de nombreux items (plus de 10) après une connexion pèse lourd dans l'ensemble des produits achetés. Le modèle choisi devra donc refléter correctement ces comportements.
L'analyse du comportement séparé des agents A et B, détaillé dans le tableau 3.12, montre que les deux types d'agent ont un comportement relativement similaire : ils achètent un ou plusieurs items. Le fait d'acheter plusieurs produits simultanément n'est donc pas spécifique aux agents A ou aux agents B.
Nous avons ensuite étudié si l'achat de plusieurs items n'était pas concentré sur certains logins. Nous avons pour cela distingué deux types de logins: ceux dont le nombre maximum d'items achetés est inférieur à 10 items, ceux dont le nombre maximum d'items achetés supérieur à 10 items.

Tableau 3.13 : Nbre items achetes après connexion
Cas avec nbre maximum d'items achetes inf à 10 items

Coun

		Type		
		A	B	Tota
Nbr item achete aprè connexio	0	971	656	1627
	1	161	78	239
	2	49	30	79
	3	19	10	29
	4	10	4	14
	5	7	4	11
	6	13		13
	7	5	1	6
	8	1		1
	9	3	1	4
Total		998	669	1667

Tableau 3.14 : Nbre items achetes après connexion (login maximum d'items achetés sup à 10 items)

Count

		Type d'agents		
		A	B	Total
Nbre items achetes après connexion	0	2858	2765	5623
	1	49	33	82
	2	11	6	17
	3	9	4	13
	4	4	1	5
	5	1		1
	6	2		2
	7	4		4
	8	1		1
	9	3		3
	10	4	2	6
	11	1	1	2
	12	1	2	3
	13	4		4
	14	2	2	4
	15	1		1
	16	1		1
	17		1	1
	20		1	1
	21	2		2
	22	1	1	2
	24	1		1
	25	1		1
	26	1		1
	64	1		1
Total		2963	2819	5782

On constate que le comportement d'achat de plusieurs items simultanément n'est pas spécifique aux agents A ou B ni à un nombre limité de logins, puisque 22 logins ont un nombre maximum d'items achetés inférieur à 10 items et 112 logins ont un nombre maximum d'items achetés supérieur à 10 items.

Pour modéliser le nombre d'items achetés après une connexion des agents A (ou B), nous avons testé plusieurs modèles existant dans la littérature en marketing.

Dans un premier temps, nous avons testé si la probabilité d'achat après une connexion suivait une loi de Poisson simple (première hypothèse du modèle NBD proposé par Morrison & Smittlein (1988)) Pour cela nous avons calculé sur excel, pour chaque login, les paramètres de la loi de Poisson modélisée à partir des données réelles de Mille Mercis. La méthode utilisée est la maximisation du maximum de vraisemblance, que nous avons calculée sous excel selon les recommandations de Hardie et Fader (2002).

Le modèle de Poisson représentant le nombre d'items achetés après une connexion est le suivant :

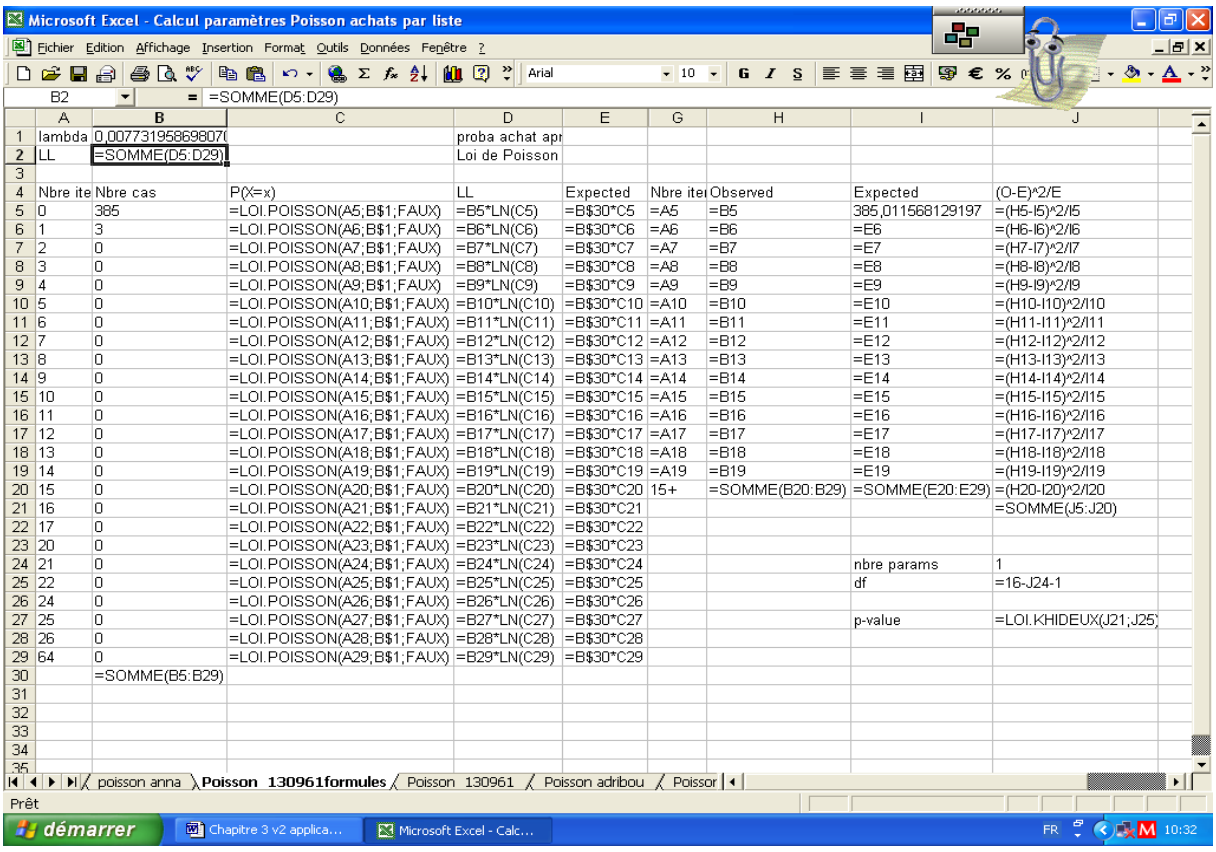
$$P(X = x|\lambda) = \lambda^x e^{-\lambda} / x!$$

X est la variable aléatoire représentant le nombre d'items achetés après une connexion donnée.

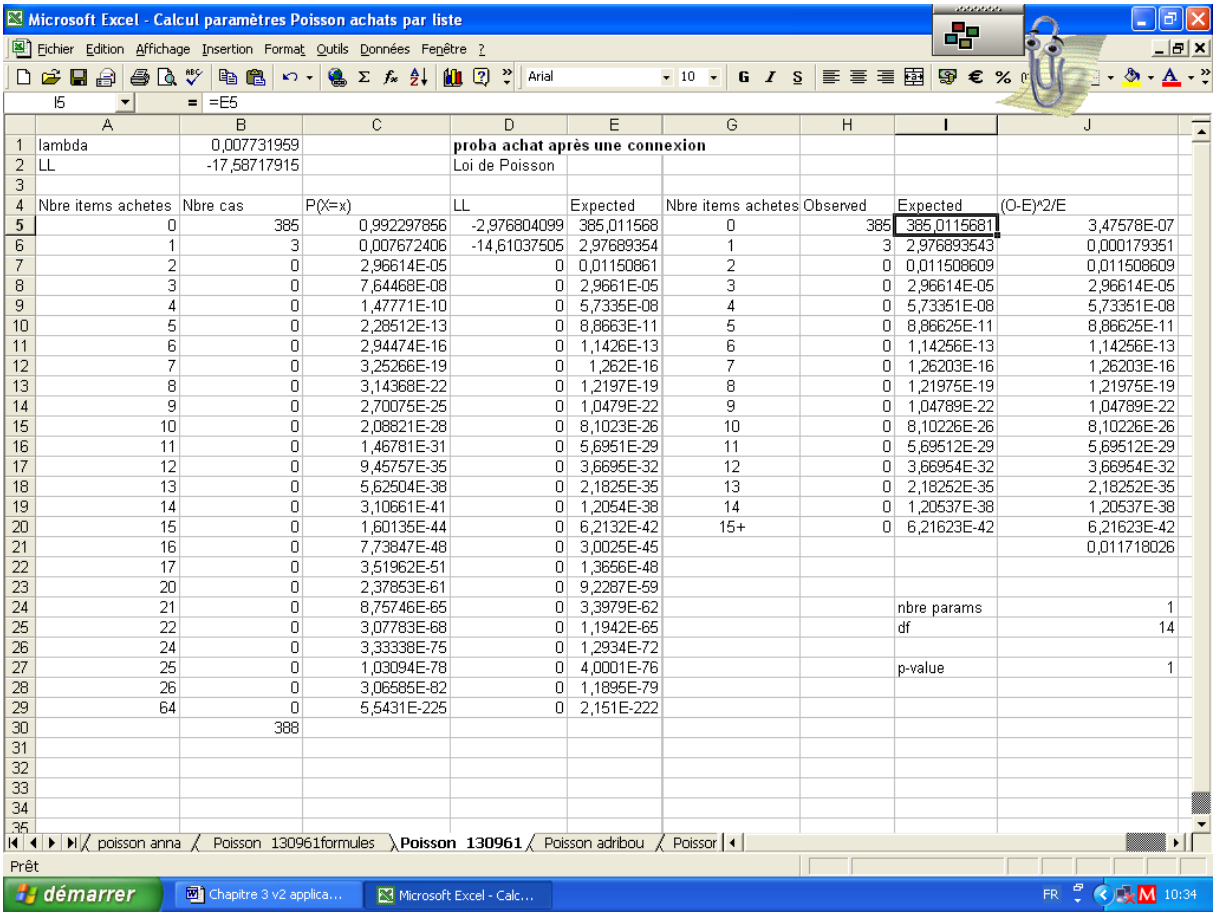
Chaque agent a donc un paramètre lambda spécifique.

Un exemple est donné ci-dessous sur l'agent A nommé 130961.

Afin de montrer le détail du modèle, nous avons présenté ci-dessous un premier tableau des formules utilisées sous excel pour calculer le maximum de vraisemblance de la loi de Poisson (cas de l'agent 130961), puis un deuxième tableau présentant les résultats.



	A	B	C	D	E	G	H	I	J
1	lambda	0,007731958698807(		proba achat apr					
2	LL	=SOMME(D5:D29)		Loi de Poisson					
3									
4	Nbre ite	Nbre cas	P(X=x)	LL	Expected	Nbre ite	Observed	Expected	(O-E)^2/E
5	0	385	=LOI.POISSON(A5;B$1;FAUX)	=B5*LN(C5)	=B$30*C5	=A5	=B5	385,011568129197	=(H5-I5)^2/I5
6	1	3	=LOI.POISSON(A6;B$1;FAUX)	=B6*LN(C6)	=B$30*C6	=A6	=B6	=E6	=(H6-I6)^2/I6
7	2	0	=LOI.POISSON(A7;B$1;FAUX)	=B7*LN(C7)	=B$30*C7	=A7	=B7	=E7	=(H7-I7)^2/I7
8	3	0	=LOI.POISSON(A8;B$1;FAUX)	=B8*LN(C8)	=B$30*C8	=A8	=B8	=E8	=(H8-I8)^2/I8
9	4	0	=LOI.POISSON(A9;B$1;FAUX)	=B9*LN(C9)	=B$30*C9	=A9	=B9	=E9	=(H9-I9)^2/I9
10	5	0	=LOI.POISSON(A10;B$1;FAUX)	=B10*LN(C10)	=B$30*C10	=A10	=B10	=E10	=(H10-I10)^2/I10
11	6	0	=LOI.POISSON(A11;B$1;FAUX)	=B11*LN(C11)	=B$30*C11	=A11	=B11	=E11	=(H11-I11)^2/I11
12	7	0	=LOI.POISSON(A12;B$1;FAUX)	=B12*LN(C12)	=B$30*C12	=A12	=B12	=E12	=(H12-I12)^2/I12
13	8	0	=LOI.POISSON(A13;B$1;FAUX)	=B13*LN(C13)	=B$30*C13	=A13	=B13	=E13	=(H13-I13)^2/I13
14	9	0	=LOI.POISSON(A14;B$1;FAUX)	=B14*LN(C14)	=B$30*C14	=A14	=B14	=E14	=(H14-I14)^2/I14
15	10	0	=LOI.POISSON(A15;B$1;FAUX)	=B15*LN(C15)	=B$30*C15	=A15	=B15	=E15	=(H15-I15)^2/I15
16	11	0	=LOI.POISSON(A16;B$1;FAUX)	=B16*LN(C16)	=B$30*C16	=A16	=B16	=E16	=(H16-I16)^2/I16
17	12	0	=LOI.POISSON(A17;B$1;FAUX)	=B17*LN(C17)	=B$30*C17	=A17	=B17	=E17	=(H17-I17)^2/I17
18	13	0	=LOI.POISSON(A18;B$1;FAUX)	=B18*LN(C18)	=B$30*C18	=A18	=B18	=E18	=(H18-I18)^2/I18
19	14	0	=LOI.POISSON(A19;B$1;FAUX)	=B19*LN(C19)	=B$30*C19	=A19	=B19	=E19	=(H19-I19)^2/I19
20	15	0	=LOI.POISSON(A20;B$1;FAUX)	=B20*LN(C20)	=B$30*C20	15+	=SOMME(B20:B29)	=SOMME(E20:E29)	=(H20-I20)^2/I20
21	16	0	=LOI.POISSON(A21;B$1;FAUX)	=B21*LN(C21)	=B$30*C21				=SOMME(J5:J20)
22	17	0	=LOI.POISSON(A22;B$1;FAUX)	=B22*LN(C22)	=B$30*C22				
23	20	0	=LOI.POISSON(A23;B$1;FAUX)	=B23*LN(C23)	=B$30*C23				
24	21	0	=LOI.POISSON(A24;B$1;FAUX)	=B24*LN(C24)	=B$30*C24			nbre params	1
25	22	0	=LOI.POISSON(A25;B$1;FAUX)	=B25*LN(C25)	=B$30*C25			df	=16-J24-1
26	24	0	=LOI.POISSON(A26;B$1;FAUX)	=B26*LN(C26)	=B$30*C26				
27	25	0	=LOI.POISSON(A27;B$1;FAUX)	=B27*LN(C27)	=B$30*C27			p-value	=LOI.KHIDEUX(J21;J25)
28	26	0	=LOI.POISSON(A28;B$1;FAUX)	=B28*LN(C28)	=B$30*C28				
29	64	0	=LOI.POISSON(A29;B$1;FAUX)	=B29*LN(C29)	=B$30*C29				
30		=SOMME(B5:B29)							
31									
32									
33									
34									
35									



Microsoft Excel - Calcul paramètres Poisson achats par liste

Fichier Edition Affichage Insertion Format Outils Données Fenêtre ?

=E5

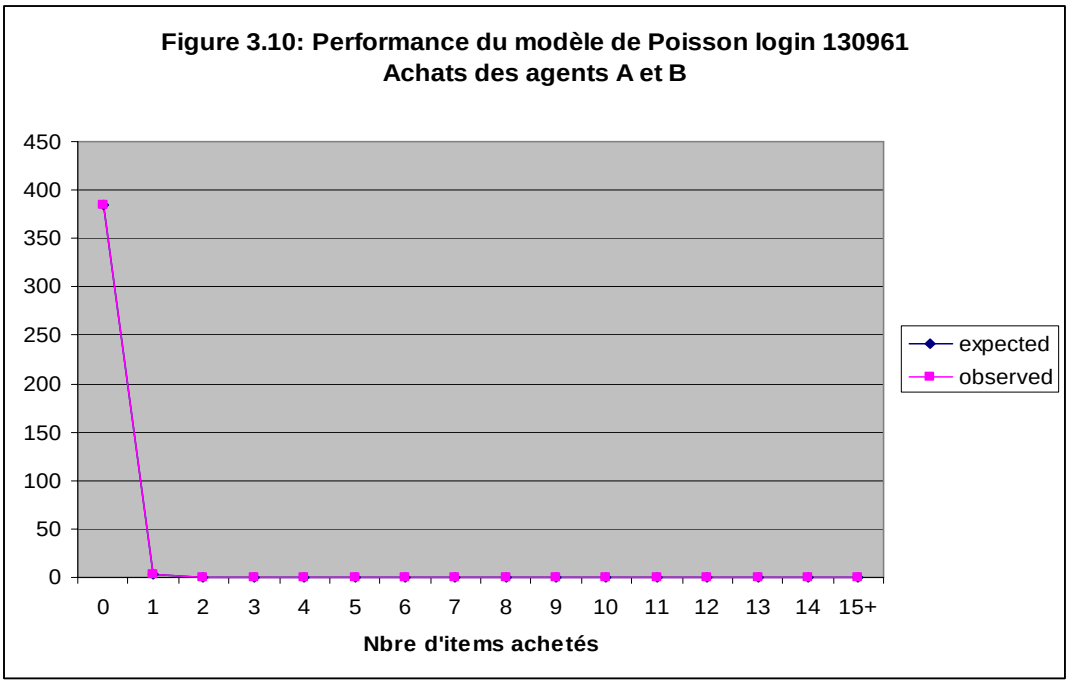
	A	B	C	D	E	G	H	I	J
1	lambda	0,007731959		proba achat après une connexion					
2	LL	-17,58717915		Loi de Poisson					
3									
4	Nbre items achetes	Nbre cas	P(X=x)	LL	Expected	Nbre items achetes	Observed	Expected	(O-E)^2/E
5	0	385	0,992297856	-2,976804099	385,011568	0	385	385,0115681	3,47578E-07
6	1	3	0,007672406	-14,61037505	2,97689354	1	3	2,976893543	0,000179351
7	2	0	2,96614E-05	0	0,01150861	2	0	0,011508609	0,011508609
8	3	0	7,64468E-08	0	2,9661E-05	3	0	2,96614E-05	2,96614E-05
9	4	0	1,47771E-10	0	5,7335E-08	4	0	5,73351E-08	5,73351E-08
10	5	0	2,28512E-13	0	8,8663E-11	5	0	8,86625E-11	8,86625E-11
11	6	0	2,94474E-16	0	1,1426E-13	6	0	1,14256E-13	1,14256E-13
12	7	0	3,25266E-19	0	1,262E-16	7	0	1,26203E-16	1,26203E-16
13	8	0	3,14368E-22	0	1,2197E-19	8	0	1,21975E-19	1,21975E-19
14	9	0	2,70075E-25	0	1,0479E-22	9	0	1,04789E-22	1,04789E-22
15	10	0	2,08821E-28	0	8,1023E-26	10	0	8,10226E-26	8,10226E-26
16	11	0	1,46781E-31	0	5,6951E-29	11	0	5,69512E-29	5,69512E-29
17	12	0	9,45757E-35	0	3,6695E-32	12	0	3,66954E-32	3,66954E-32
18	13	0	5,62504E-38	0	2,1825E-35	13	0	2,18252E-35	2,18252E-35
19	14	0	3,10661E-41	0	1,2054E-38	14	0	1,20537E-38	1,20537E-38
20	15	0	1,60135E-44	0	6,2132E-42	15+	0	6,21623E-42	6,21623E-42
21	16	0	7,73847E-48	0	3,0025E-45				0,011718026
22	17	0	3,51962E-51	0	1,3656E-48				
23	20	0	2,37853E-61	0	9,2287E-59				
24	21	0	8,75746E-65	0	3,3979E-62			nbre params	1
25	22	0	3,07783E-68	0	1,1942E-65			df	14
26	24	0	3,33338E-75	0	1,2934E-72				
27	25	0	1,03094E-78	0	4,0001E-76			p-value	1
28	26	0	3,06585E-82	0	1,1895E-79				
29	64	0	5,5431E-225	0	2,151E-222				
30		388							
31									
32									
33									
34									
35									

poisson anna | Poisson 130961formules | **Poisson 130961** | Poisson adribou | Poissor

Prêt

démarrer | Chapitre 3 v2 applica... | Microsoft Excel - Calc... | FR | 10:34

La performance du modèle de Poisson est représentée dans la figure 3.10 , dans laquelle les données du modèle correspondent parfaitement aux données observées.



On constate donc que sur cet agent 130961 les achats suivent effectivement une loi de Poisson de paramètre lambda=0,0077. Le test du Khi-deux donne un niveau de signification p-value égal à 1 (supérieur à 0,05), ce qui nous permet de conclure que les achats après connexion (de A ou de B) pour cet agent 130961 suivent une loi de Poisson.

Ce même travail a été mené sur tous les agents A. Le tableau 3.11 ci-dessous montre l'ensemble des paramètres lambda et les niveaux de signification du Khi-deux pour chacun des 135 agents.

Tableau 3.11 : résultats de la loi de Poisson par login

Résultats Loi de poisson		
agentA	Lambda	P-value
130961	0,00773	1
adribou	0,17073	0
akwell1	0,18644	0
alexO	0,05136	0
ama1	0,07143	1
amibio	0,04082	0,02994
anakeen	0,25	1
anna	0,0216	0
apopo	0,1	1
ashley85	0,01695	1
aude	0,10687	0
badir	0,07143	0
bgougay	0,03665	0,94672
blondi	0,03333	1
bontze	0,02273	1
bpdl	0,10227	0
brusson1	0,01393	0
bruyere	0,01961	1
bshunke	0,11364	0
carolala	0,01493	1
cathou1	0,04348	1
cathou25	0,38462	0,65496
cclerget	0,3	0,99989
celine21	0,05797	0,88417
chap's	0,15152	0
chrispi1	0,09302	0,988
clemisa	0,06	0,67962
cypher	0,27083	0
dalmate	0,2439	0
delatte	0,07006	0,85858
delille	0,10256	0,99314
didine77	0,2	0
dinou	0,14607	0,01741
djoule	0,14686	0
dolores	0,04167	1
doppel	0,5	0
dperrot	0,003	1
dzhouari	0,0625	1
eb95	0,28889	0,00531
eb96	0,4	0
ehran	0,24	0
emilie	0,24	0
estia	0,09756	0
eusebio	0,075	0,84041
fab007	0,14286	0,88932
fabby93	0,11628	0,99958
fabou	0,07407	0,41316
falouis	0,13793	0,96301
fdrugeat	0,33333	1
fevrard	0,07143	1
fournial	0,21212	0
fruor	0,07692	1
frzpower	0,07407	0,41316
gaelleM	0,26667	0,99962
gallouin	0,03001	0
gilamon	1	1
ginie31	0,00635	0
gougay	0,04255	0,03984
greensky	0,15476	0
guix	0,01042	0
gzing	0,06306	0
hedi.c	0,03704	1
herpin	0,14	0
hutch	0,06154	0,00174
j-kolb	0,36364	0
jeanbask	0,34783	0
jeannot3	0,02913	0
jeda	0,03704	1
jimasa	1,3	0
joewj	0,05405	0,14671
josetta	1,09091	0
juboubou	0,02857	0,00104
juju75	0,03846	0,01925
kabotine	0,08906	0
kathrin	0,05747	0
kifran	0,14474	0
lbarbier	0,0396	0
leleu5	0,76923	0
levieux	0,0566	0,98723
lilimimi	0,04	1
locquet	0,0678	0
loicT	0,43902	0
luberdov	0,69231	0
ludovicb	0,26923	0
m@hy	0,03247	0
manwe	0,25714	0
mary_ce	0,21951	0
mayfair	0,001	1
mekiela	0,28571	0
michel	0,07843	1
micka74	0,14706	0
mike_s	0,04854	0
miloumoi	0,37037	0
minious	0,06731	0,99881
modjo13	0,41176	0
mperrot	0,0989	0
munier	0,10135	0
naemesis	0,59813	0
nickus	0,06667	0
ninie12	0,35	0
noel_net	0,30612	0
philia	0,02273	0,02273
picardj	0,001	1
popcorn	0,05063	0
poupide	0,16667	0,9347
prioton	0,02128	1
prozaik	0,04651	0
quentin	0,08333	0
raph	0,07692	0
relindis	0,27273	0
rlaval	0,00939	0
romain37	0,02778	1
sandor	0,03571	1
sebulba9	0,05	1
semery	0,02655	0
sicart	0,05263	1
sophynet	0,22857	0
souris	0,02381	1
starchen	0,8125	0

	Lambda	P-value
starchen	0,8125	0
steph	0,04706	0,00005
stumpf	0,3871	0
sylvie4	0,02703	1
tantan	4,33333	0
thibault	0,03339	0
tonio2k	0,01316	1
toto2	0,05	0,10163
tranb	0,01266	1
vianney2	0,02303	0
virginie	0,01	1
xzouf	0,10465	0
yseulis	0,08	0
yves	0,04255	0,03984

La synthèse des résultats des niveaux de signification p-value du test du Khi-deux est la suivante :

Tableau 3.12 Nombre de cas où les achats suivent une loi de Poisson (test du khi-deux significatif)

	pvalue sup à 0,05	pvalue sup à 0,03	pvalue sup à 0,01
	Count	Count	Count
0	80	78	74
1	54	56	60

Ce tableau montre que, dans le cas où alpha=5%, les achats après connexion suivent une loi de Poisson pour 54 agents A, mais que pour 80 agents A on ne peut conclure à une modélisation correcte par une loi de Poisson.

En étant moins strict sur le test du khi-deux, on obtient pour alpha=1% une loi de Poisson dans uniquement 60 cas sur 135, ce qui reste insuffisant pour conclure à une telle loi modélisant les données observées.

Nous avons donc ensuite cherché à améliorer le modèle de Poisson.

Pour cela, nous avons introduit dans le modèle une hétérogénéité, pour chaque login, au sein des connexions (un paramètre lambda par connexion), ce qui nous a naturellement conduit à étudier un modèle NBD (Negative Binomial Distribution), modèle classique en marketing pour représenter les achats des consommateurs (Morrisson & Schmittlein, 1988).

Le modèle NBD est défini dans notre cas de la manière suivante:

Soit X le nombre d'items achetés après une connexion,

pour chaque connexion, on fait l'hypothèse que X suit une loi de Poisson de paramètre λ:

$$P(X = x|\lambda) = \lambda^x e^{-\lambda} / x!$$

Les paramètres (λ) sont distribués sur les différentes connexions selon une loi Gamma :

$$g(\lambda) = \alpha^r \lambda^{r-1} e^{-\alpha\lambda} / \Gamma(r)$$

Pour calculer les valeurs du modèle pour chaque connexion, on applique sur excel les équations suivantes, recommandées par Hardie et Foster (2001):

$$P(X=x) = (\alpha / (\alpha+1))^r \quad \text{pour } x=0$$

$$P(X=x) = ((r+x-1)/ x(\alpha+1)) \; P(X = x-1) \text{ pour } x \geq 1$$

Puis on estime les paramètres du modèle par la méthode du maximum de vraisemblance en utilisant le solveur excel.

Afin de détailler le modèle utilisé, les tableaux suivants montrent un exemple sur le login AnnA. Le tableau 3.12 présente les formules utilisées sous excel pour calculer les paramètres du modèle, le tableau 3.13 montre les résultats obtenus.

.

Tableau 3.12 : Modèle NBD sur le login AnnA (Formules)

r	0,041935124626677!								
alpha	1,94113236227179								
pi									
LL	=SOMME(E8:E32)								
Nbre items achetes	Nbre cas	P(X=x) NBD	LL	Expected		Nbre items achet	Observed	Expected	(O-E)^2/E
0	2047	=(B2/(B2+1))^B1	=B8*LN(C8)	=B$33*C8		=A8	=B8	=F8	=(I8-J8)^2/J8
1	30	=(B$1+A9-1)/(A9*(B$2+1))*C8	=B9*LN(C9)	=B$33*C9		=A9	=B9	=F9	=(I9-J9)^2/J9
2	4	=(B$1+A10-1)/(A10*(B$2+1))*C9	=B10*LN(C10)	=B$33*C10		=A10	=B10	=F10	=(I10-J10)^2/J10
3	1	=(B$1+A11-1)/(A11*(B$2+1))*C10	=B11*LN(C11)	=B$33*C11		=A11	=B11	=F11	=(I11-J11)^2/J11
4	1	=(B$1+A12-1)/(A12*(B$2+1))*C11	=B12*LN(C12)	=B$33*C12		=A12	=B12	=F12	=(I12-J12)^2/J12
5	0	=(B$1+A13-1)/(A13*(B$2+1))*C12	=B13*LN(C13)	=B$33*C13		=A13	=B13	=F13	=(I13-J13)^2/J13
6	0	=(B$1+A14-1)/(A14*(B$2+1))*C13	=B14*LN(C14)	=B$33*C14		=A14	=B14	=F14	=(I14-J14)^2/J14
7	0	=(B$1+A15-1)/(A15*(B$2+1))*C14	=B15*LN(C15)	=B$33*C15		=A15	=B15	=F15	=(I15-J15)^2/J15
8	0	=(B$1+A16-1)/(A16*(B$2+1))*C15	=B16*LN(C16)	=B$33*C16		=A16	=B16	=F16	=(I16-J16)^2/J16
9	0	=(B$1+A17-1)/(A17*(B$2+1))*C16	=B17*LN(C17)	=B$33*C17		=A17	=B17	=F17	=(I17-J17)^2/J17
10	0	=(B$1+A18-1)/(A18*(B$2+1))*C17	=B18*LN(C18)	=B$33*C18		=A18	=B18	=F18	=(I18-J18)^2/J18
11	0	=(B$1+A19-1)/(A19*(B$2+1))*C18	=B19*LN(C19)	=B$33*C19		=A19	=B19	=F19	=(I19-J19)^2/J19
12	0	=(B$1+A20-1)/(A20*(B$2+1))*C19	=B20*LN(C20)	=B$33*C20		=A20	=B20	=F20	=(I20-J20)^2/J20
13	0	=(B$1+A21-1)/(A21*(B$2+1))*C20	=B21*LN(C21)	=B$33*C21		=A21	=B21	=F21	=(I21-J21)^2/J21
14	0	=(B$1+A22-1)/(A22*(B$2+1))*C21	=B22*LN(C22)	=B$33*C22		=A22	=B22	=F22	=(I22-J22)^2/J22
15	0	=(B$1+A23-1)/(A23*(B$2+1))*C22	=B23*LN(C23)	=B$33*C23		15+	=SOMME(B23:B32)	=SOMME(F23:F32)	=(I23-J23)^2/J23
16	0	=(B$1+A24-1)/(A24*(B$2+1))*C23	=B24*LN(C24)	=B$33*C24					=SOMME(K8:K23)
17	0	=(B$1+A25-1)/(A25*(B$2+1))*C24	=B25*LN(C25)	=B$33*C25					
20	0	=(B$1+A26-1)/(A26*(B$2+1))*C25	=B26*LN(C26)	=B$33*C26					
21	0	=(B$1+A27-1)/(A27*(B$2+1))*C26	=B27*LN(C27)	=B$33*C27				nbre params	2
22	0	=(B$1+A28-1)/(A28*(B$2+1))*C27	=B28*LN(C28)	=B$33*C28				df	=16-K27-1
24	0	=(B$1+A29-1)/(A29*(B$2+1))*C28	=B29*LN(C29)	=B$33*C29					
25	0	=(B$1+A30-1)/(A30*(B$2+1))*C29	=B30*LN(C30)	=B$33*C30				pvalue	=LOI.KHIDEUX(K24;K28)
26	0	=(B$1+A31-1)/(A31*(B$2+1))*C30	=B31*LN(C31)	=B$33*C31					
64	0	=(B$1+A32-1)/(A32*(B$2+1))*C31	=B32*LN(C32)	=B$33*C32					
	=SOMME(B8:B32)								

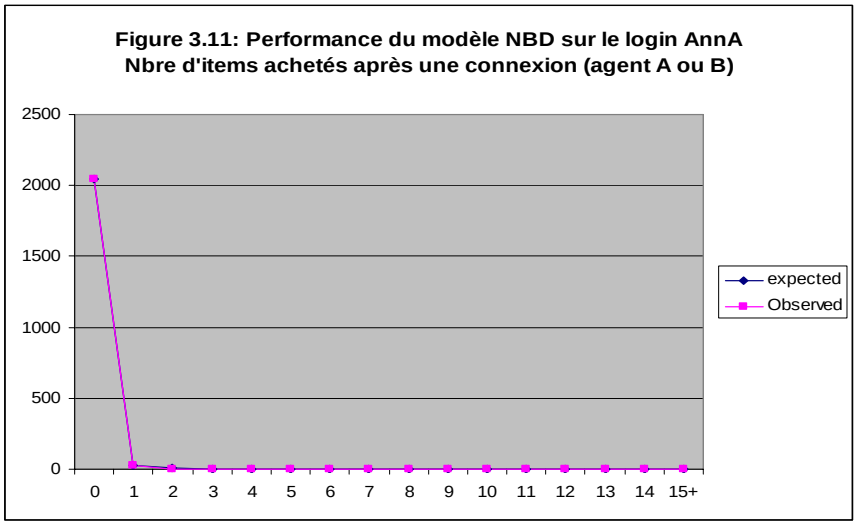
Tableau 3.13 : Modèle NBD sur le login AnnA (Résultats)

r	0,041935125	**proba achat après une connexion**	
alpha	1,941132362		
pi			
LL	-203,9764364		

Nbre items ac	Nbre cas	P(X=x) NBD	LL	Expected
0	2047	0,9827259	-35,6690076	2047,01809
1	30	0,0140119	-128,035536	29,1867037
2	4	0,0024819	-23,9948534	5,16988834
3	1	0,0005744	-7,46222236	1,19642996
4	1	0,0001485	-8,81481751	0,30935894
5	0	4,082E-05	0	0,08502907
6	0	1,166E-05	0	0,02429399
7	0	3,423E-06	0	0,00712955
8	0	1,024E-06	0	0,00213378
9	0	3,112E-07	0	0,00064827
10	0	9,568E-08	0	0,0001993
11	0	2,97E-08	0	6,186E-05
12	0	9,291E-09	0	1,9354E-05
13	0	2,926E-09	0	6,0953E-06
14	0	9,268E-10	0	1,9306E-06
15	0	2,95E-10	0	6,145E-07
16	0	9,43E-11	0	1,9642E-07
17	0	3,025E-11	0	6,302E-08
20	0	9,794E-12	0	2,0401E-08
21	0	3,178E-12	0	6,6199E-09
22	0	1,034E-12	0	2,1528E-09
24	0	3,374E-13	0	7,0274E-10
25	0	1,103E-13	0	2,2978E-10
26	0	3,612E-14	0	7,5247E-11
64	0	1,21E-14	0	2,5201E-11
	2083			

Nbre items achetes	Observed	Expected	(O-E)^2/E
0	2047	2047,01809	1,6E-07
1	30	29,1867037	0,0226627
2	4	5,16988834	0,2647327
3	1	1,19642996	0,0322499
4	1	0,30935894	1,54185
5	0	0,08502907	0,0850291
6	0	0,02429399	0,024294
7	0	0,00712955	0,0071296
8	0	0,00213378	0,0021338
9	0	0,00064827	0,0006483
10	0	0,0001993	0,0001993
11	0	6,186E-05	6,186E-05
12	0	1,9354E-05	1,935E-05
13	0	6,0953E-06	6,095E-06
14	0	1,9306E-06	1,931E-06
15+	0	9,0414E-07	9,041E-07
			1,9810196
		nbre params	2
		df	13
		pvalue	0,9997856

Le cas d'AnnA ci-dessus a été volontairement sélectionné en raison du nombre important de données (2083 connexions d'agents A ou B) et du comportement d'achat (plusieurs items sont parfois achetés après une connexion). Le tableau 3.13 montre de bons résultats du modèle NBD pour le cas AnnA (p-value= 0.99), ce qui est reflété par la figure 3.11 ci-dessous, qui compare les données du modèle avec celles observées (les deux courbes sont superposées).



Etant donné le nombre important de connexions suivies d'aucun achat, nous avons également testé le Zero Inflated Negative Binomial Distribution, qui affecte un poids aux données qui sont à 0.

Le modèle est le suivant :

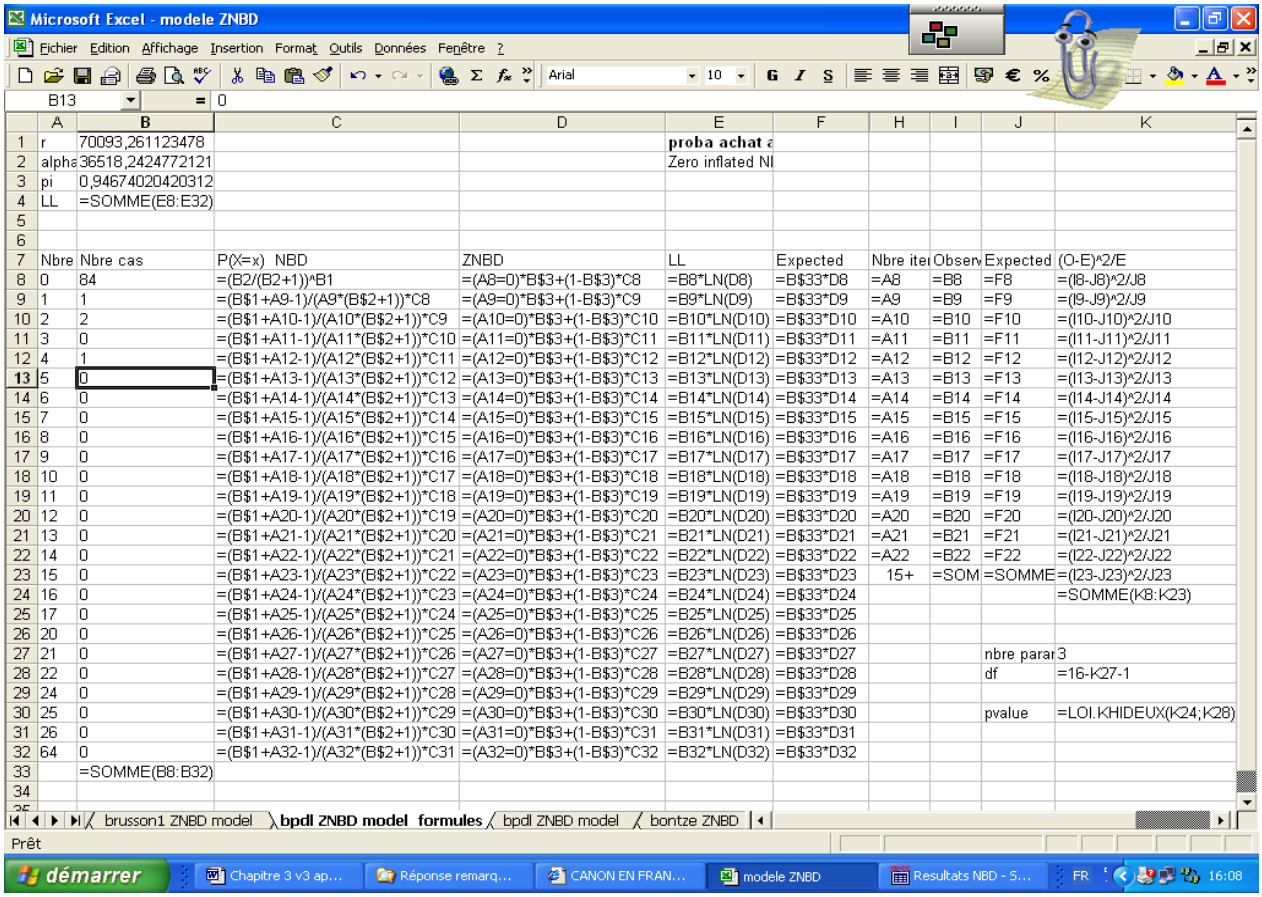
Si π est la proportion de connexions suivies de zéro achat, on a :

$$P(X=x) = \delta_{x=0}\,\pi + (1-\pi)\;\left(\Gamma(r+x)/\,\Gamma(r)\,x!\right)\left(\alpha/\alpha+1\right)^{r}\left(1/\alpha+1\right)^{x}$$

Ce modèle a été appliqué par login. L'estimation des paramètres s'effectue sous excel par la méthode du maximum de vraisemblance (Hardie et Fader, 20002).

A titre d'exemple, nous illustrons dans les tableaux ci-dessous le modèle ZNBD sur le login bpdl. Le choix de ce login provient du fait que le nombre d'items achetés pour ce login n'était pas bien modélisé par une loi de Poisson. Le tableau 3.14 donne les formules utilisées, le tableau 3.15 présente les résultats sur ce login.

Tableau 3.14 : Modèle ZNBD sur le login bpdl (Formules)



Microsoft Excel - modele ZNBD

Fichier Edition Affichage Insertion Format Outils Données Fenêtre ?

B13 = 0

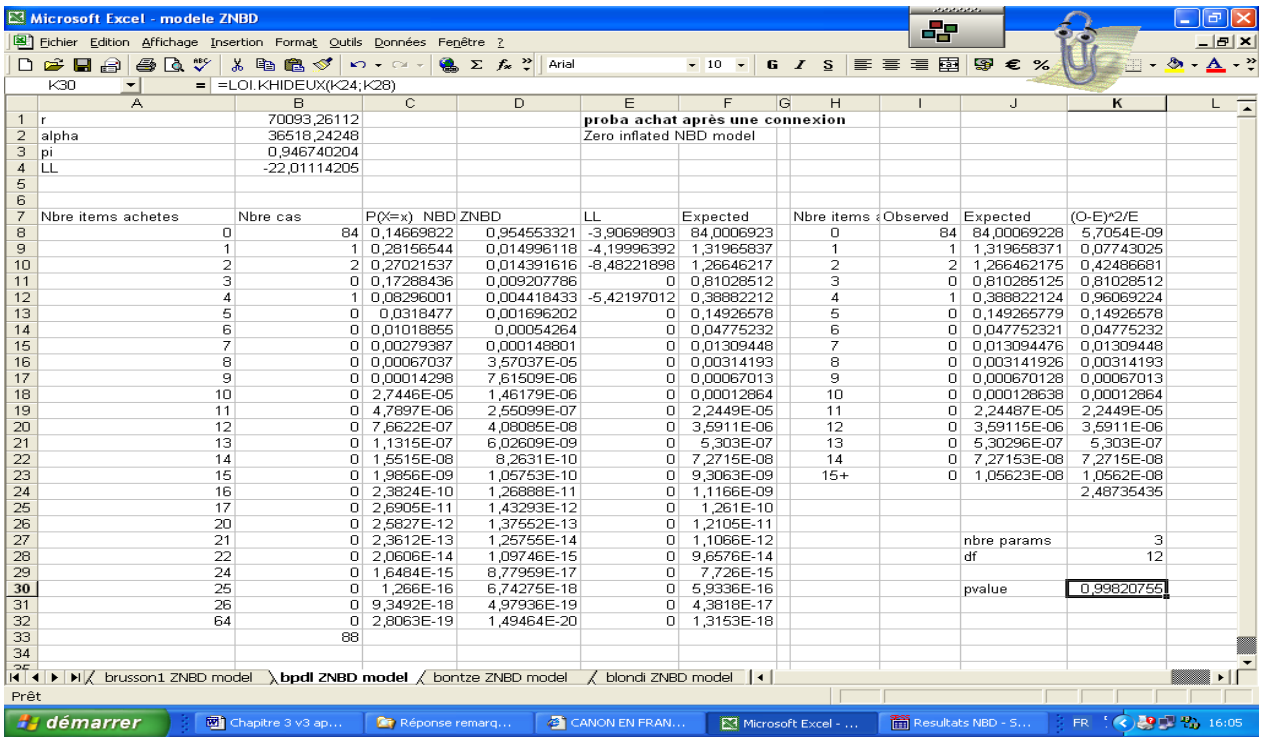
	A	B	C	D	E	F	H	I	J	K
1	r	70093,261123478			proba achat a					
2	alpha	36518,2424772121			Zero inflated NI					
3	pi	0,94674020420312								
4	LL	=SOMME(E8:E32)								
5										
6										
7	Nbre	Nbre cas	P(X=x) NBD	ZNBD	LL	Expected	Nbre iter	Observ	Expected	(O-E)^2/E
8	0	84	=(B2/(B2+1))^B1	=(A8=0)*B$3+(1-B$3)*C8	=B8*LN(D8)	=B$33*D8	=A8	=B8	=F8	=(I8-J8)^2/J8
9	1	1	=(B$1+A9-1)/(A9*(B$2+1))*C8	=(A9=0)*B$3+(1-B$3)*C9	=B9*LN(D9)	=B$33*D9	=A9	=B9	=F9	=(I9-J9)^2/J9
10	2	2	=(B$1+A10-1)/(A10*(B$2+1))*C9	=(A10=0)*B$3+(1-B$3)*C10	=B10*LN(D10)	=B$33*D10	=A10	=B10	=F10	=(I10-J10)^2/J10
11	3	0	=(B$1+A11-1)/(A11*(B$2+1))*C10	=(A11=0)*B$3+(1-B$3)*C11	=B11*LN(D11)	=B$33*D11	=A11	=B11	=F11	=(I11-J11)^2/J11
12	4	1	=(B$1+A12-1)/(A12*(B$2+1))*C11	=(A12=0)*B$3+(1-B$3)*C12	=B12*LN(D12)	=B$33*D12	=A12	=B12	=F12	=(I12-J12)^2/J12
13	5	0	=(B$1+A13-1)/(A13*(B$2+1))*C12	=(A13=0)*B$3+(1-B$3)*C13	=B13*LN(D13)	=B$33*D13	=A13	=B13	=F13	=(I13-J13)^2/J13
14	6	0	=(B$1+A14-1)/(A14*(B$2+1))*C13	=(A14=0)*B$3+(1-B$3)*C14	=B14*LN(D14)	=B$33*D14	=A14	=B14	=F14	=(I14-J14)^2/J14
15	7	0	=(B$1+A15-1)/(A15*(B$2+1))*C14	=(A15=0)*B$3+(1-B$3)*C15	=B15*LN(D15)	=B$33*D15	=A15	=B15	=F15	=(I15-J15)^2/J15
16	8	0	=(B$1+A16-1)/(A16*(B$2+1))*C15	=(A16=0)*B$3+(1-B$3)*C16	=B16*LN(D16)	=B$33*D16	=A16	=B16	=F16	=(I16-J16)^2/J16
17	9	0	=(B$1+A17-1)/(A17*(B$2+1))*C16	=(A17=0)*B$3+(1-B$3)*C17	=B17*LN(D17)	=B$33*D17	=A17	=B17	=F17	=(I17-J17)^2/J17
18	10	0	=(B$1+A18-1)/(A18*(B$2+1))*C17	=(A18=0)*B$3+(1-B$3)*C18	=B18*LN(D18)	=B$33*D18	=A18	=B18	=F18	=(I18-J18)^2/J18
19	11	0	=(B$1+A19-1)/(A19*(B$2+1))*C18	=(A19=0)*B$3+(1-B$3)*C19	=B19*LN(D19)	=B$33*D19	=A19	=B19	=F19	=(I19-J19)^2/J19
20	12	0	=(B$1+A20-1)/(A20*(B$2+1))*C19	=(A20=0)*B$3+(1-B$3)*C20	=B20*LN(D20)	=B$33*D20	=A20	=B20	=F20	=(I20-J20)^2/J20
21	13	0	=(B$1+A21-1)/(A21*(B$2+1))*C20	=(A21=0)*B$3+(1-B$3)*C21	=B21*LN(D21)	=B$33*D21	=A21	=B21	=F21	=(I21-J21)^2/J21
22	14	0	=(B$1+A22-1)/(A22*(B$2+1))*C21	=(A22=0)*B$3+(1-B$3)*C22	=B22*LN(D22)	=B$33*D22	=A22	=B22	=F22	=(I22-J22)^2/J22
23	15	0	=(B$1+A23-1)/(A23*(B$2+1))*C22	=(A23=0)*B$3+(1-B$3)*C23	=B23*LN(D23)	=B$33*D23	15+	=SOM	=SOMME	=(I23-J23)^2/J23
24	16	0	=(B$1+A24-1)/(A24*(B$2+1))*C23	=(A24=0)*B$3+(1-B$3)*C24	=B24*LN(D24)	=B$33*D24				=SOMME(K8:K23)
25	17	0	=(B$1+A25-1)/(A25*(B$2+1))*C24	=(A25=0)*B$3+(1-B$3)*C25	=B25*LN(D25)	=B$33*D25				
26	20	0	=(B$1+A26-1)/(A26*(B$2+1))*C25	=(A26=0)*B$3+(1-B$3)*C26	=B26*LN(D26)	=B$33*D26				
27	21	0	=(B$1+A27-1)/(A27*(B$2+1))*C26	=(A27=0)*B$3+(1-B$3)*C27	=B27*LN(D27)	=B$33*D27			nbre parar	3
28	22	0	=(B$1+A28-1)/(A28*(B$2+1))*C27	=(A28=0)*B$3+(1-B$3)*C28	=B28*LN(D28)	=B$33*D28			df	=16-K27-1
29	24	0	=(B$1+A29-1)/(A29*(B$2+1))*C28	=(A29=0)*B$3+(1-B$3)*C29	=B29*LN(D29)	=B$33*D29				
30	25	0	=(B$1+A30-1)/(A30*(B$2+1))*C29	=(A30=0)*B$3+(1-B$3)*C30	=B30*LN(D30)	=B$33*D30			pvalue	=LOI.KHIDEUX(K24;K28)
31	26	0	=(B$1+A31-1)/(A31*(B$2+1))*C30	=(A31=0)*B$3+(1-B$3)*C31	=B31*LN(D31)	=B$33*D31				
32	64	0	=(B$1+A32-1)/(A32*(B$2+1))*C31	=(A32=0)*B$3+(1-B$3)*C32	=B32*LN(D32)	=B$33*D32				
33		=SOMME(B8:B32)								
34										

brusson1 ZNBD model | bpdl ZNBD model formules | bpdl ZNBD model | bontze ZNBD

Prêt

démarrer | Chapitre 3 v3 ap... | Réponse remarq... | CANON EN FRAN... | modele ZNBD | Resultats NBD - S... | FR | 16:08

Tableau 3.15 : Modèle ZNBD sur le login bpdl (Résultats)



Microsoft Excel - modele ZNBD

Fichier Edition Affichage Insertion Format Outils Données Fenêtre ?

K30 = =LOI.KHIDEUX(K24;K28)

	A	B	C	D	E	F	G	H	I	J	K	L
1	r	70093,26112			proba achat après une connexion							
2	alpha	36518,24248			Zero inflated NBD model							
3	pi	0,946740204										
4	LL	-22,01114205										
5												
6												
7	Nbre items achetes	Nbre cas	P(X=x) NBD	ZNBD	LL	Expected		Nbre items	Observed	Expected	(O-E)^2/E	
8	0	84	0,14669822	0,954553321	-3,90698903	84,0006923		0	84	84,00069228	5,7054E-09	
9	1	1	0,28156544	0,014996118	-4,19996392	1,31965837		1	1	1,319658371	0,07743025	
10	2	2	0,27021537	0,014391616	-8,48221898	1,26646217		2	2	1,266462175	0,42486681	
11	3	0	0,17288436	0,009207786	0	0,81028512		3	0	0,810285125	0,81028512	
12	4	1	0,08296001	0,004418433	-5,42197012	0,38882212		4	1	0,388822124	0,96069224	
13	5	0	0,0318477	0,001696202	0	0,14926578		5	0	0,149265779	0,14926578	
14	6	0	0,01018855	0,00054264	0	0,04775232		6	0	0,047752321	0,04775232	
15	7	0	0,00279387	0,000148801	0	0,01309448		7	0	0,013094476	0,01309448	
16	8	0	0,00067037	3,57037E-05	0	0,00314193		8	0	0,003141926	0,00314193	
17	9	0	0,00014298	7,61509E-06	0	0,00067013		9	0	0,000670128	0,00067013	
18	10	0	2,7446E-05	1,46179E-06	0	0,00012864		10	0	0,000128638	0,00012864	
19	11	0	4,7897E-06	2,55099E-07	0	2,2449E-05		11	0	2,24487E-05	2,2449E-05	
20	12	0	7,6622E-07	4,08085E-08	0	3,5911E-06		12	0	3,59115E-06	3,5911E-06	
21	13	0	1,1315E-07	6,02609E-09	0	5,303E-07		13	0	5,30296E-07	5,303E-07	
22	14	0	1,5515E-08	8,2631E-10	0	7,2715E-08		14	0	7,27153E-08	7,2715E-08	
23	15	0	1,9856E-09	1,05753E-10	0	9,3063E-09		15+	0	1,05623E-08	1,0562E-08	
24	16	0	2,3824E-10	1,26888E-11	0	1,1166E-09					2,48735435	
25	17	0	2,6905E-11	1,43293E-12	0	1,261E-10						
26	20	0	2,5827E-12	1,37552E-13	0	1,2105E-11						
27	21	0	2,3612E-13	1,25755E-14	0	1,1066E-12				nbre params	3	
28	22	0	2,0606E-14	1,0974E-15	0	9,6576E-14				df	12	
29	24	0	1,6484E-15	8,77959E-17	0	7,726E-15						
30	25	0	1,266E-16	6,74275E-18	0	5,9336E-16				pvalue	0,99820755	
31	26	0	9,3492E-18	4,97936E-19	0	4,3818E-17						
32	64	0	2,8063E-19	1,49464E-20	0	1,3153E-18						
33		88										
34												

brusson1 ZNBD model | bpdl ZNBD model | bontze ZNBD model | blondi ZNBD model

Prêt

démarrer | Chapitre 3 v3 ap... | Réponse remarq... | CANON EN FRAN... | Microsoft Excel - ... | Resultats NBD - S... | FR | 16:05

On constate pour ce login qu'on obtient un résultat du test du chi-deux significatif (0,99) donc un modèle Zero Inflated NBD qui représente correctement les données réelles, avec trois paramètres r (70093) , α (36518) , et π (0,946).

Nous avons mené une analyse similaire sur l'ensemble des logins, puis effectué un test du ratio de vraisemblance sur chacun des login afin de déterminer l'apport du modèle ZNBD par rapport au modèle NBD simple. Les résultats comparés des modèles NBD et ZNBD sont fournis dans le tableau 3.16.

Principe du test de ratio de la vraisemblance :

Soit 2 modèles A et B. Si en fixant k contraintes sur les paramètres du modèle A, on obtient le modèle B, on dit que le modèle B est emboîté dans le modèle A. Par exemple, le modèle NBD est emboîté dans le modèle ZNBD, puisqu'en fixant π=0 dans le modèle ZNBD, on obtient le modèle NBD. Le test du ratio de vraisemblance détermine si le modèle A, avec plus de paramètres, est meilleur que le modèle B.

Le ratio de vraisemblance, noté LR, est calculé de la manière suivante :

$$LR = -2 (LL_B - LL_A)$$

On teste ensuite l'hypothèse nulle : le modèle A n'est pas différent du modèle B

La règle de décision est le rejet de l'hypothèse nulle si $LR > \chi^2_{.05,k}$

Si on applique ce test dans le cadre de nos données, la première conclusion issue du tableau 3.16 est que le modèle NBD est meilleur que le modèle ZNBD dans 85,8% des cas (115 cas sur 135).

La deuxième conclusion tirée du tableau 3.16 est que le modèle NBD représente correctement les données observées (à la différence des modèles de Poisson étudiés précédemment).Cette affirmation provient des résultast du test du Khi-2 (colonne p-value NBD du tableau 3.16) : le modèle NBD ne converge pas sur excell dans uniquement 3 cas sur 135, alors que le modèle ZNBD ne converge pas dans 20 cas. De plus, les p-value sont supérieures à 0.05 et souvent égales à 1 pour le modèle NBD, ce qui n'est pas le cas pour le modèle ZNBD.

Nous retiendrons donc de ces analyses que le comportement d'achat des agents A et B après une connexion sur un même login du site www.MilleMercis.com suit une loi NBD à deux paramètres, et ce pour chaque login étudié.

Tableau 3.16: comparaison des modèles

Login	LL ZNBD	ZNBD NCP*	ZNBD p value	LL NBD	NBD pvalue	LL ratio test	chi deux	NBD best
130961	-17,654	0	1	-17,59	1	-0,13	3,84	1
adribou	-16,2288	0	0,699	-15,34	0,83	-1,79	3,84	1
akwell1	-72,349	0	0,007	-26,6	0,09	-91,49	3,84	1
AlexO	-64,5367	0	0,836	-19,87	1	-89,34	3,84	1
ama1	-7,2057	0	1	-3,94	1	-6,52	3,84	1
amibio	-6,0092	0	0,999	0	1	-12,02	3,84	1
anakeen	-12,4329	0	1	-12,62	1	0,38	3,84	1
AnnA	-205,096	0	0,81	-96,57	0,99	-217,06	3,84	1
apopo	-8,3896	0	0,973	-5,08	1	-6,62	3,84	1
ashley85	-5,0952	0	1	-5,01	1	-0,17	3,84	1
Aude	-26,2259	0	0,049	-12,62	0,48	-27,2	3,84	1
badir	-6,6306	0	0,983	0	1	-13,26	3,84	1
bgougay	-29,4859	0	1	-25,82	1	-7,33	3,84	1
blondi	-4,4105	0	1	-4,31	1	-0,21	3,84	1
bontze	-4,7731	0	1	-4,78	1	0,02	3,84	1
bpdl	-22,0111	0	0,998	-22,91	0,96	1,79	3,84	1
brusson1	-30,5718	0	0	-9,83	0,16	-41,48	3,84	1
bruyere	-4,9481	0	1	0	1	-9,9	3,84	1
bschunke	-14,3169	0	0,997	-12,68	1	-3,28	3,84	1
carolala	-5,2234	0	1	0	1	-10,45	3,84	1
cathou1	-4,1396	0	1	-0,58	1	-7,13	3,84	1
CATHOU25	-10,3875	0	0,997	-10,39	1	0,01	3,84	1
cclerget	-6,7693	0	1	-2,42	1	-8,7	3,84	1
Céline21	-14,4893	0	1	-7,14	1	-14,71	3,84	1
chap's	-24,8453	0	0,991	-9,7	1	-30,29	3,84	1
Chrispi1	-13,0294	0	1	-13,2	1	0,34	3,84	1
clemisa	-10,1623	0	1	-10,42	1	0,52	3,84	1
Cypher	-24,4939	0	0,729	-2,42	1	-44,15	3,84	1
DALMATE	-6,7798	0	0,859	0	1	-13,56	3,84	1
Delatte	-39,5485	0	0,999	-11,25	1	-56,59	3,84	1
delillle	-12,7254	0	1	-12,89	1	0,33	3,84	1
didine77	-46,4901	0	0	0	1	-92,98	3,84	1
Dinou	-37,0609	0	1	-11	1	-52,13	3,84	1
djoule	-466,94	0	0,062	-110,87	0,01	-712,15	3,84	1
dolores3	-4,1831	0	1	-4,14	1	-0,1	3,84	1
doppel	-19,3539	0	0,964	-15,71	0,96	-7,3	3,84	1
dperrot	.	1	.	.	NCP*	.	3,84	.
dzhouari	-3,7668	0	1	-3,77	1	0,01	3,84	1
eb95	-27,8638	0	1	-13,25	0,99	-29,22	3,84	1
eb96	-9,7077	0	0,832	-3,56	1	-12,29	3,84	1
ehran	-32,3646	0	0,295	-58,22	0	51,72	3,84	.
Emilie	-18,5826	0	0	0	1	-37,17	3,84	1
estia	-10,6558	0	0,999	-6,13	0,84	-9,05	3,84	1
eusebio1	-9,7259	0	1	-9,96	1	0,47	3,84	1
Fab007	-4,7366	0	0,999	-6,58	0,84	3,7	3,84	1
fabby93	-15,7324	0	1	-16,45	1	1,44	3,84	1
fabou	-5,4047	0	0,999	0	1	-10,81	3,84	1

Login	LL ZNBD	ZNBD NCP*	ZNBD p valu	LL NBD	NBD pvalue	LL ratio test	chi deux	NBD meilleur
fallouis	-23,5914	0	1	-15,85	1	-15,48	3,84	1
fdrugeat	-3,8213	0	1	-4,2	1	0,75	3,84	1
fevrard	-7,2057	0	1	-3,86	1	-6,68	3,84	1
FOURNIAL	-33,7918	0	0,946	-15,61	1	-36,36	3,84	1
Fruor	-3,5516	0	1	-3,62	1	0,13	3,84	1
frzpower	-5,4047	0	0,999	-5,86	0,99	0,9	3,84	1
GaëlleM	-19,3416	0	1	-12,35	0,99	-13,99	3,84	1
gallouin	-63,3303	0	0,636	-49,26	0,6	-28,13	3,84	1
gilamon	.	1	1	.	NCP*	.	3,84	.
Ginie31	-8,341	1	0,985	-8,34	0,99	0	3,84	1
gougay	-7,8856	0	0,066	-7,17	0,78	-1,44	3,84	1
Greensky	-27,4089	0	0,727	-18,3	0,96	-18,22	3,84	1
guix	.	1	.	-7,84	0,99	.	3,84	.
gzing	-17,6765	0	0	0	1	-35,35	3,84	1
hedi.dc	-12,8324	0	1	-12,95	1	0,23	3,84	1
HERPIN	.	1	.	0	1	.	3,84	.
hutch	-11,1866	0	0,98	-6,42	0,99	-9,52	3,84	1
j-kolb	-4,9717	0	0,982	-11,22	0	12,51	3,84	.
jeanbask	-14,1608	0	0,971	-3,49	1	-21,35	3,84	1
jeannot3	-163,657	0	0	-91,28	0,64	-144,75	3,84	1
jeda	-4,3032	0	1	0	1	-8,61	3,84	1
JIMASA	-5,4587	0	0,779	0	1	-10,92	3,84	1
joewj	-5,725	0	0,999	0	1	-11,45	3,84	1
josetta	-48,729	0	0	-7,41	0	-82,63	3,84	1
juboubou	.	1	.	0	1	.	3,84	.
juju75	-6,0692	0	0,999	0	1	-12,14	3,84	1
kabotine	-103,718	0	0,991	-31,08	0,99	-145,28	3,84	1
Kathrin	-20,3935	0	0	-8,41	0,16	-23,97	3,84	1
Kifran	-103,092	0	0	0	1	-206,18	3,84	1
lbarbier	.	1	.	0	1	.	3,84	.
leleu5	-5,6051	0	0,857	-8,4	0,09	5,59	3,84	.
levieux	-22,7909	0	1	0	1	-45,58	3,84	1
lilimimi	-0,015	1	1	.	NCP*	.	3,84	.
LoïcT	-81,6781	0	0	-17,75	0,94	-127,86	3,84	1
Loquet	-22,4043	0	1	0	1	-44,81	3,84	1
luberdov	-5,5548	0	0,884	0	1	-11,11	3,84	1
ludovicb	-11,2169	0	0,568	-11,24	0,49	0,04	3,84	1
m@hy	.	1	.	0	1	.	3,84	.
manwe	-18,6804	0	0	0	1	-37,36	3,84	1
mary_ce	-18,0451	0	0	0	1	-36,09	3,84	1
Mayfair	-0,0388	1	1	0	1	-0,08	3,84	1
mekiela	-18,5289	0	0,2	-10,49	0,99	-16,08	3,84	1
Michel	-14,023	0	1	0	1	-28,05	3,84	1
micka74	-10,3217	0	0,999	-11,43	0,97	2,22	3,84	1
mike_s	-12,2825	0	0	-8,58	0,16	-7,41	3,84	1
miloumoi	-64,2101	0	0	-8,13	0	-112,17	3,84	1
minious	-25,7574	0	1	0	1	-51,51	3,84	1

Login	LL ZNBD	ZNBD NCP*	ZNBD p valu	LL NBD	NBD pvalue	LL ratio test	chi deux	NBD meilleur
modjo-13	-14,2762	0	0	-7,19	0,01	-14,18	3,84	1
mperrot	-88,7423	0	0	-11,89	1	-153,7	3,84	1
Munier	.	1	.	0	1	.	3,84	.
Naemesis	.	1	.	0	1	.	3,84	.
nhai07	.	1	.	0	1	.	3,84	.
nickus	.	1	.	-7,02	0,83	.	3,84	.
ninie12	-5,9102	0	0,918	0	1	-11,82	3,84	1
noel_net	-13,9461	0	0	-13,95	0	0	3,84	1
philia	-17,654	0	1	-5,76	1	-23,79	3,84	1
picardjf	-0,0341	1	1	0	1	-0,07	3,84	1
popcorn	-22,3649	0	0	0	1	-44,73	3,84	1
poupide	-4,5707	0	0,999	-5,01	0,99	0,87	3,84	1
prioton	-4,8656	0	1	0	1	-9,73	3,84	1
prozaik	-40,9946	0	0	0	1	-81,99	3,84	1
quentin	-43,6656	0	0	0	1	-87,33	3,84	1
RAPH	-9,2173	0	0	0	1	-18,43	3,84	1
relindis	-5,8959	0	0,95	-7,25	0,04	2,71	3,84	1
rlaval	.	1	.	-20,57	1	.	3,84	.
romain37	-4,5957	0	1	0	1	-9,19	3,84	1
sandor	-4,3403	0	1	0	1	-8,68	3,84	1
sebulba9	-3,9965	0	1	-4	1	0	3,84	1
semery	-206,506	0	0	-67,21	1	-278,6	3,84	1
sicart	-3,9438	0	1	-3,96	1	0,04	3,84	1
sophynet	-45,3668	0	0	0	1	-90,73	3,84	1
Souris	-4,7518	0	1	0	1	-9,5	3,84	1
starchen	-29,4048	0	0	0	1	-58,81	3,84	1
Steph	-11,7305	0	0,98	0	1	-23,46	3,84	1
stumpf	.	1	.	0	1	.	3,84	.
Sylvie4	-4,6234	0	1	0	1	-9,25	3,84	1
Tantan	-105,702	0	0	-7,51	0,82	-196,38	3,84	1
Thibaut	-142,558	0	0,044	-30,04	0,84	-225,03	3,84	1
Tonio2k	-5,3503	0	1	-5,33	1	-0,04	3,84	1
toto2	-5,8039	0	0,999	0	1	-11,61	3,84	1
tranb	-5,3893	0	1	0	1	-10,78	3,84	1
vianney2	.	1	.	-22,1	0,96	.	3,84	.
Virginie	-0,0722	1	1	0	1	-0,14	3,84	1
xzouf	-122,691	0	0	-6,2	1	-232,98	3,84	1
Yseulys	1,91E+10	0	0	-41,73	0,94	3,81E+10	3,84	.
Yves	-5,9723	0	0,999	0	1	-11,94	3,84	1
zaultar	-0,0001	1	1	0	1	0	3,84	1

Légende du tableau **:**

LLZNBD : maximum de vraisemblance avec le modèle ZNBD

*ZNBD NCP** : le modèle ZNBD ne converge pas sur excell

ZNBD p-value : p-value du test du chi-deux appliqué au modèle ZNBD

LLNBD : maximum de vraisemblance avec le modèle NBD

NBD p-value : p-value du test du chi-deux appliqué au modèle NBD (les cas avec NCP* correspondent aux cas où le modèle NBD ne converge pas sur excell)

Llratiotest : valeur du ratio test du maximum de vraisemblance permettant de comparer les deux modèles NBD et ZNBD.

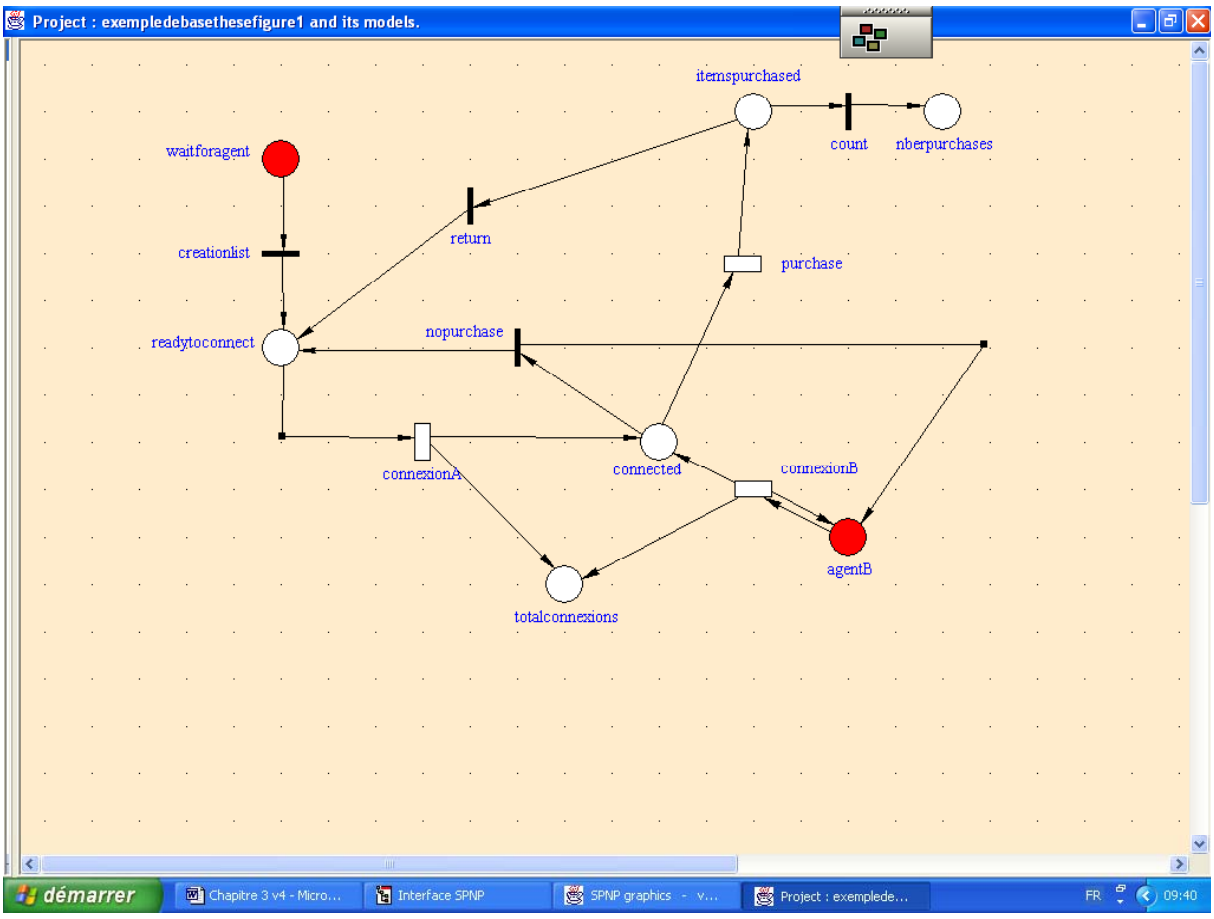
Chideux : valeur du chi deux à un degré de liberté (permet de faire le LL ratiotest)

NBD meilleur : égal à 1 lorsque le LL ratio test est inférieur au Khi-2 à un degré de liberté (alors le modèle ZNBD est moins bon que le modèle NBD).

3.4.4 Intégration des lois stochastiques dans le modèle de Petri :

L'étape suivante de notre démarche est d'intégrer les différentes lois stochastiques (loi de connexion des agents A, loi de connexion des agents B, loi d'achat) dans un système global : le réseau de Petri. Rappelons ici la figure 3.2 du réseau de Petri qui constitue le modèle de base d'une wish list :

Figure 3.2 : Wish List représentée par un réseau de Petri (modèle de base)



Les transitions connexionA, connexionB et purchase sont donc paramétrées dans le modèle Petri (logiciel SPNP) selon les lois précédemment identifiées, avec leurs paramètres correspondants.
Les résultats du modèle, sa performance comparée à d'autres modèles plus « classiques » en marketing ainsi qu'une extension du modèle de base seront présentés dans le chapitre 4.

CONCLUSION :

L'objectif de ce chapitre était de modéliser un système de wish list par la méthodologie des réseaux de Petri, ce que nous avons réalisé à partir de données empiriques issues du site Web de la société Millemercis dont l'activité principale est de gérer des wish lists. Nous avons dans ce modèle décomposé les wish lists en un certain nombre d'actions séquentielles menées par les internautes (connexions, achat). Nous avons tout d'abord essayé de modéliser chacune de ces actions et finalement trouvé des lois stochastiques pour chaque action (Weibull pour les connexions et NBD pour le comportement d'achat). Nous avons également mis en avant la possibilité de modéliser un système multi-agents par les réseaux de Petri, puisqu'une wish list est bien un SMA. Le modèle que nous avons développé intègre les actions de plusieurs agents (A ou B). Nous mettons aussi en avant l'un des avantages des réseaux de Petri qui est de représenter des actions parallèles : notre modèle par exemple reflète bien la situation d'un agent A se connectant sur le site pendant qu'un agent B achète en parallèle des cadeaux sur la liste de l'agent A. De plus, nous montrons que les réseaux de Petri stochastiques sont des modèles permettant d'intégrer plusieurs lois stochastiques modélisant différents événements du système.
Dans le chapitre qui suit, nous mesurons la performance de notre modèle en intégrant les paramètres des loi de Weibull (connexion agent A, connexion agent B) et NBD (comportement d'achat) dans le réseau de Petri de la figure 3.2. Afin de montrer que les réseaux de Petri constituent une approche innovante et performante par rapport à d'autres modèles plus « classiques » en marketing, nous effectuons également dans le chapitre suivant une comparaison de notre modèle avec ces approches plus classiques. Enfin, nous développons un modèle plus complexe, intégrant des actions marketing (comme l'e-mailing) et effectuons ensuite des simulations afin de montrer un autre avantage des réseaux de Petri qui est de pouvoir simuler différentes stratégies marketing, de les comparer, et d'aider les managers à prendre leur décision en terme d'actions marketing à mener.

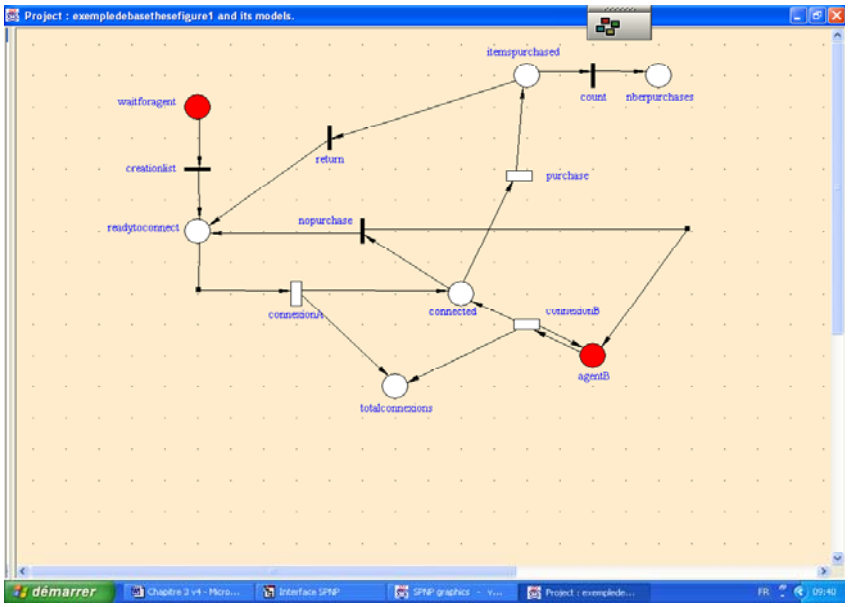
Chapitre 4
Les réseaux de Petri pour modéliser les systèmes de wish list : performance du modèle et extension à la mesure de l'impact de l'e-mailing

4.1 INTRODUCTION :

Nous avons dans le chapitre précédent présenté le principe des réseaux de Petri permettant de modéliser les wish lists. Nous avons ensuite détaillé la modélisation des lois décrivant chacun des évènements stochastiques du système, à savoir : la connexion des agents A sur le site www.MilleMercis.com, celle des agents B sur ce même site et enfin le comportement d'achat de ces agents A et B après une connexion. Les autres transitions sont traitées dans le modèle comme des évènements immédiats et ne suivent donc pas de loi spécifique. Dans le chapitre 4, nous allons tout d'abord présenter les résultats de ce modèle de Petri appliqué aux wish lists. Nous allons ensuite étudier les performances de ce type de modèle, en le comparant notamment à un modèle plus classiquement utilisé en marketing (le modèle NBD avec régression). Nous développerons enfin une extension du modèle pour mesurer l'impact de l'e-mailing.

4.2 RESULTATS DES RESEAUX DE PETRI POUR MODELISER LES WISH LISTS :

Rappelons tout d'abord le modèle utilisé pour représenter les wish lists, décrit précisément dans le chapitre précédent (figure 3.2 du chapitre 3) :



4.2.1 Sélection des données pour le calcul des paramètres des lois stochastiques :

Etant donné que la date de création d'une liste varie d'un agent à l'autre, nous avons sélectionné une période de temps permettant d'une part de calibrer notre modèle et d'autre part d'effectuer des simulations et des prédictions. L'analyse des données disponibles sur les logins nous a permis de déterminer une durée maximale de 12 mois pour la calibration du modèle et de 4 mois pour la prédiction. Nous avons donc exclu les logins qui n'avaient pas suffisamment d'historique (moins de 16 mois d'historique) et avons travaillé au final sur 94 agents A. La première étape consista par conséquent à déterminer les paramètres des lois des transitions *connexionA, connexionB* et *purchase* décrits dans le réseau de Petri ci-dessus (Figure 3.2 du chapitre 3)

Loi de la durée interconnexion des agents A :

En appliquant les résultats du chapitre 3 précédent, nous avons calculé sur excel par la méthode du maximum de vraisemblance les paramètres de la loi de Weibull représentant la durée d'interconnexion des agents A sur la période de calibration du modèle, à savoir les 12 mois suivant la date de création de la première liste pour chaque agent A.

Les résultats obtenus pour les paramètres de Weibull sont fournis dans le tableau 4.1 ci-dessous. Ce tableau comprend les paramètres calculés directement sur excel par la méthode du maximum de vraisemblance, puis les paramètres de la loi de Weibull recalculés pour les intégrer dans le modèle de Petri, car les formules décrivant la loi de Weibull diffèrent de celles utilisées dans excel. La différence de paramètres entre Weibull dans les réseaux de Petri et dans excel est la suivante :

val1 (paramètre alpha1) dans les réseaux de Petri = (1/ Béta excel)$^{\text{alpha excel}}$

val2 (paramètre béta1) dans les réseaux de Petri = alpha excel

En effet, les formules pour la fonction de densité de probabilité de la loi de Weibull pour chaque cas sont les suivantes :

Excel : $f(x;\alpha;\beta)=\frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}e^{-(x/\beta)^{\alpha}}$ Equation 1

Petri : $f(x;val1;val2)=val1\,val2\,x^{val2-1}e^{-val1*x^{val2}}$ Equation 2

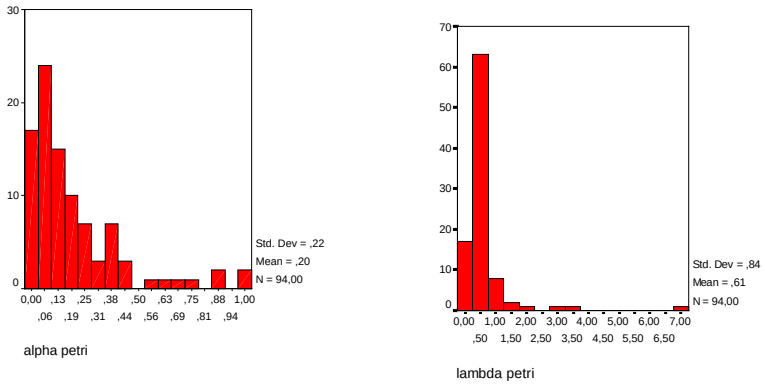
Tableau 4.1:paramètres de la loi de Weibull

login	alpha weibull excel	beta weibull excel	alpha Pétri	lambda Pétri
130961	0,5000	42,8478	0,1528	0,5000
akwell1	0,5355	93,2256	0,0882	0,5355
AlexO	0,4822	148,7919	0,0896	0,4822
ama1	0,1677	168,7369	0,4231	0,1677
amibio	0,2926	83,4709	0,2740	0,2926
AnnA	0,4401	3,8068	0,5553	0,4401
Aude	0,3114	703,1734	0,1298	0,3114
badir	0,2755	10438,4591	0,0781	0,2755
bgougay	0,5547	42,6163	0,1247	0,5547
blondi	0,2312	53,9772	0,3976	0,2312
bontze	0,6226	66,2906	0,0734	0,6226
bpdl	0,5072	121,8228	0,0875	0,5072
bruyere	0,5019	786,5686	0,0352	0,5019
bschunke	0,6804	459,0447	0,0155	0,6804
carolala	0,6054	89,9534	0,0656	0,6054
cathou1	0,0100	2,0000	0,9931	0,0100
cclerget	0,0300	98,0000	0,8715	0,0300
Céline21	0,9242	219,0229	0,0069	0,9242
chap's	0,1976	300,0359	0,3240	0,1976
Cypher	0,3734	1089,9143	0,0734	0,3734
DALMATE	0,4266	230,4038	0,0982	0,4266
Delatte	0,5415	123,7626	0,0736	0,5415
didine77	0,6821	105,7867	0,0416	0,6821
Dinou	0,3607	176,9285	0,1546	0,3607
djoule	0,6300	6,5751	0,3053	0,6300
dolores3	0,3235	2211,6260	0,0828	0,3235
doppel	0,2925	8496,4100	0,0709	0,2925
eb95	0,2364	141,7629	0,3101	0,2364
eb96	0,1579	210294,7582	0,1445	0,1579
Emilie	0,3789	86,4513	0,1846	0,3789
estia	0,5261	198,4331	0,0618	0,5261
eusebio1	0,2909	308,3407	0,1887	0,2909
fabou	1,3946	320,2976	0,0003	1,3946
fallouis	0,3472	238,0863	0,1496	0,3472
fevrard	0,2854	2110,3525	0,1125	0,2854
FOURNIAL	1,1788	407,6091	0,0008	1,1788
Fruor	0,1345	2,3372	0,8921	0,1345
frzpower	0,2031	2218,0348	0,2091	0,2031
GaëlleM	0,3684	471,6083	0,1036	0,3684
gallouin	0,4483	102,0342	0,1258	0,4483
Ginie31	0,7253	1512,5783	0,0049	0,7253
Greensky	0,2858	2225,6693	0,1105	0,2858
gzing	0,7667	225,4542	0,0157	0,7667
HERPIN	0,4547	79,8859	0,1365	0,4547
hutch	0,6314	142,8851	0,0436	0,6314
jeanbask	1,1788	407,6089	0,0008	1,1788
jeannot3	0,5049	7,1183	0,3712	0,5049
jeda	0,0800	345,0000	0,6266	0,0800

login	alpha weibull excell	beta weibull excell	alpha Pétri	lambda Pétri
JIMASA	0,2814	673,9739	0,1600	0,2814
joewj	0,4642	249,9708	0,0771	0,4642
juboubou	0,4428	305,8066	0,0793	0,4428
juju75	0,4528	41,9449	0,1842	0,4528
kabotine	0,6166	48,3772	0,0915	0,6166
Kifran	0,5675	182,0028	0,0522	0,5675
lbarbier	0,4581	22,5974	0,2397	0,4581
levieux	0,5084	416,8969	0,0466	0,5084
Loquet	1,9234	219,5459	0,0000	1,9234
luberdov	0,0202	1000000,0000	0,7561	0,0202
m@hy	0,2962	29,8366	0,3658	0,2962
manwe	0,6365	120,1227	0,0475	0,6365
mary_ce	0,1554	211,6247	0,4351	0,1554
Mayfair	0,1821	247,1012	0,3666	0,1821
mekiela	0,7038	86,0588	0,0435	0,7038
Michel	0,3395	564,5211	0,1164	0,3395
minious	1,2717	404,3928	0,0005	1,2717
mperrot	0,4982	58,0521	0,1322	0,4982
Munier	0,3231	136,9938	0,2040	0,3231
Naemesis	0,3233	116,2625	0,2149	0,3233
nhai07	0,4045	74,4346	0,1749	0,4045
ninie12	0,0614	2360844,6942	0,4062	0,0614
philia	0,3278	13,9338	0,4217	0,3278
picardjf	1,0922	757,2591	0,0007	1,0922
popcorn	0,2633	140,2869	0,2720	0,2633
prioton	0,9000	345,0000	0,0052	0,9000
prozaiik	0,3894	515,5675	0,0879	0,3894
quentin	7,0000	999,0000	0,0000	7,0000
RAPH	0,5845	63,8597	0,0881	0,5845
rlaval	0,2849	1,0034	0,9990	0,2849
romain37	0,8091	504,6232	0,0065	0,8091
sandor	0,9726	358,6805	0,0033	0,9726
semery	0,6010	8,4492	0,2773	0,6010
sophynet	0,2232	4,4382	0,7170	0,2232
Souris	0,3074	78,5799	0,2615	0,3074
starchen	3,4574	283,2408	0,0000	3,4574
Stèph	0,4873	41,3919	0,1630	0,4873
stumpf	0,7153	113,0738	0,0340	0,7153
Sylvie4	0,2804	162,0332	0,2401	0,2804
Thibaut	0,6585	15,3882	0,1653	0,6585
toto2	0,0203	0,0000	0,0000	0,0203
tranb	0,1692	3231,9044	0,2549	0,1692
vianney2	0,3142	2,8608	0,7188	0,3142
Virginie	0,5892	110,6009	0,0625	0,5892
xzouf	3,0000	98765,0000	0,0000	3,0000
Yseulys	0,6234	26,4054	0,1299	0,6234
Yves	0,7148	2086,2081	0,0042	0,7148

On constate sur ce tableau que les paramètres ne sont pas les mêmes d'un individu à l'autre, ce qui signifie que chaque agent A a un comportement de connexion propre sur le site qui suit une loi de Weibull avec des paramètres spécifiques.

Figure 4.1: Histogramme des paramètres alpha et lambda de la loi de Weibull (durée entre 2 connexions des agents A)



La figure 4.1 représente les histogrammes des paramètres alpha (Petri) et lambda (Petri) de la loi de Weibull modélisant la durée entre 2 connexions des agents A.

Nous avons dans un premier temps étudié l'existence éventuelle d'une corrélation entre les deux paramètres alpha et lambda. Nous avons mené pour cela une régression linéaire avec comme variable dépendante alpha et comme variable indépendante lambda. Les résultats de cette régression sont les suivants :

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	,254	,027		9,431	,000
	lambda petri	-9,38E-02	,026	-,350	-3,589	,001

a. Dependent Variable: alpha petri

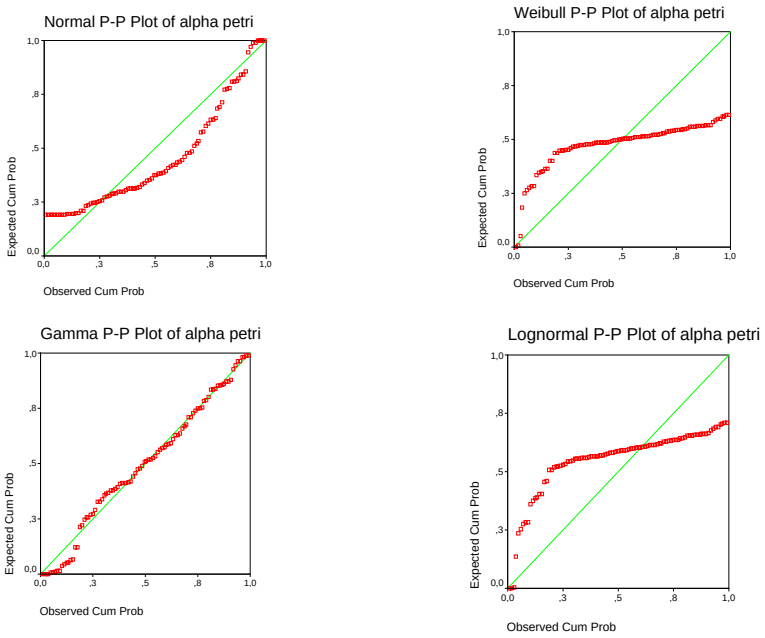
Ce tableau montre que les deux coefficients sont corrélés de manière significative (Sig.< 0.05). On pourrait donc écrire que :

$$\alpha = - 0.0938 \ \lambda + 0.254$$

En revanche, le R^2 de la régression est égal à 0.123, ce qui signifie que les deux paramètres sont corrélés de manière modérée.

Nous avons donc cherché si chacun de ces paramètres suivait une loi particulière. Parmi les lois testées (Log-Normale, Weibull, Gamma, étant donnée la forme des histogrammes ci-dessus), les lois les plus proches sont les suivantes :

Figure 4.2 : PP Plots paramètre alpha (Petri) de la loi de Weibull



La figure 4.2 montre dans un premier temps la distribution du paramètre alpha de la loi de Weibull. La distribution de ce paramètre sur l'ensemble des agents A considérés suit une loi Gamma, comme on peut le constater dans la figure 4.2, les autres lois donnant des écarts importants avec les données observées.

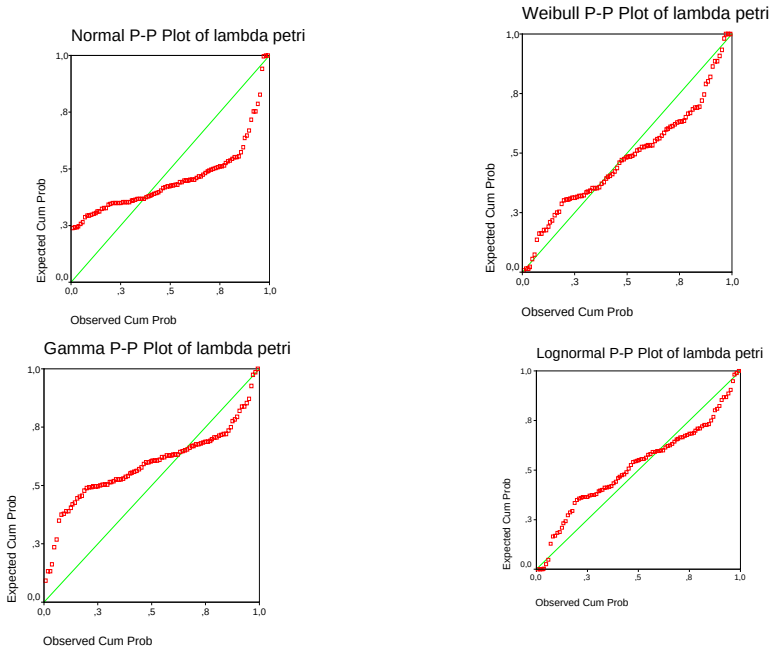
Afin de valider cette hypothèse, nous avons mené un test de Kolmogorov-Smirnov dont les résultats sont présentés ci-dessous . Ce test donne un résultat non significatif (Asymp.Sig.>0.05) nous permettant de ne pas rejeter l'hypothèse nulle et de conclure par conséquent que les deux distributions (données observées d'alpha et loi de Gamma) ne sont pas différentes.

Two sample Kolmogorov-Smirnov Test

		alpha petri
Most Extreme Differences	Absolute	,170
	Positive	,074
	Negative	-,170
Kolmogorov-Smirnov Z		1,167
Asymp. Sig. (2-tailed)		,131

On conclut donc que le paramètre alpha (Petri) suit une loi Gamma sur l'ensemble des agents A.

Figure 4.3 : PP Plots paramètre lambda de la loi de Weibull



En ce qui concerne le deuxième paramètre lambda (Petri) de la loi de Weibull, on remarque dans la figure 4.3 que la distribution de ce paramètre sur l'ensemble des agents A suit une loi de Weibull, la loi log-normale représentant aussi correctement les données observées (nous avons également analysé les QQ plots qui confirment cette affirmation). Ces deux lois donnent quasiment la même vraisemblance (-273.92 pour Weibull et – 274.54 pour Log-Normale) mais nous avons sélectionné la loi de Weibull pour des raisons de cohérence avec la loi Gamma choisie pour le paramètre alpha.

Pour confirmer que la distribution de lambda correspond à une loi de Weibull, nous avons effectué un test de Kolmogorov-Smirnov qui donne un résultat non significatif

(Asymp.Sig.>0.05) nous permettant de ne pas rejeter l'hypothèse nulle et de conclure que les deux distributions (données observées du lambda et loi de Weibull) ne sont pas différentes. Les résultats de ce test sont présentés dans le tableau ci-dessous.

Two sample Kolmogorov-Smirnov Test

		lambda petri
Most Extreme Differences	Absolute	,202
	Positive	,106
	Negative	-,202
Kolmogorov-Smirnov Z		1,386
Asymp. Sig. (2-tailed)		,053

On pourrait donc représenter la durée entre deux connexions d'un agent A sur le site Millemercis par une loi de Weibull de paramètre alpha et lambda, avec :

- alpha suit une loi Gamma de paramètres (formules adaptées au réseau de Petri) 0.8265 et 3.9766
- lambda suit une loi de Weibull de paramètre (formules adaptées au réseau de Petri) alpha 1 = 2.6099 et béta1=1.6180

Les paramètres ci-dessus ont été calculés sous excel en maximisant la vraisemblance et en utilisant la fonction solveur.

Loi de la durée interconnexion des agents B :

De la même manière, en utilisant les résultats du chapitre 3, nous avons calculé sur excel par la méthode du maximum de vraisemblance les paramètres de la loi Weibull représentant la durée d'interconnexion des agents B sur la période de calibration du modèle, à savoir les 12 mois suivant la date de création de la première liste pour chaque agent A.

Les résultats obtenus pour les paramètres de la loi de Weibull sont fournis pour chaque agent dans le tableau 4.2 ci-dessous. La loi de Weibull est définie par deux paramètres λ (scale) et c (shape), selon les définitions fournies au chapitre précédent.

Pour certains des agents A, il n'y a aucune connexion d'agent B, ce qui explique le nombre de lignes plus faible du tableau 4.2 (81 agents A sur les 94 ont des agents B qui se connectent). Ce cas correspond à la situation où des agents A ont créé leur liste sans avoir

fourni d'adresses e-mail d'amis ou de connaissances. La liste n'est alors que consultée par l'agent A qui l'a créée et sert souvent de liste d'achats.

Tableau 4.2: Paramètres de la loi de Weibull (durée interconnexion agents B)

Login	alpha (excel)	beta (excel)	shape(Pétri)	scale(Pétri)
130961	0,3598	7,0523	0,3598	0,4952
akwell1	0,3032	31,9361	0,3032	0,3499
AlexO	0,5047	32,5905	0,5047	0,1723
ama1	0,1394	643,3083	0,1394	0,4059
amibio	0,1716	4,4975	0,1716	0,7726
AnnA	0,3378	21,7403	0,3378	0,3534
Aude	0,3591	28,7432	0,3591	0,2994
badir	0,2657	21,0039	0,2657	0,4453
bgougay	0,3000	3425,0000	0,3000	0,0870
blondi	0,4583	219,4484	0,4583	0,0845
bontze	0,1968	12,1554	0,1968	0,6116
bpdl	0,1526	7,9816	0,1526	0,7284
bruyere	0,7964	80,3718	0,7964	0,0304
carolala	0,5611	170,3495	0,5611	0,0560
cathou1	0,2959	24,9026	0,2959	0,3862
Céline21	0,4762	51,7134	0,4762	0,1527
chap's	0,2500	876,0000	0,2500	0,1838
Cypher	0,5260	477,0169	0,5260	0,0390
DALMATE	0,1554	2015,8432	0,1554	0,3065
Delatte	0,2237	76,9005	0,2237	0,3785
didine77	0,2000	210,0000	0,2000	0,3432
Dinou	0,3990	57,6707	0,3990	0,1983
djoule	0,3860	6,3391	0,3860	0,4902
dolores3	4,0000	9876,0000	4,0000	0,0000
Emilie	0,1000	0,0000	0,1000	3,7145
eusebio1	0,2301	470,9622	0,2301	0,2427
fabou	0,2983	127,4725	0,2983	0,2355
fallouis	0,4997	166,1855	0,4997	0,0777
fevrard	0,2729	1003,7686	0,2729	0,1516
FOURNIAL	0,4132	14,5091	0,4132	0,3311
Fruor	0,1383	3213,6609	0,1383	0,3272
frzpower	0,5289	353,6265	0,5289	0,0449
GaëlleM	0,0950	560,4651	0,0950	0,5481
gallouin	0,2525	1,7313	0,2525	0,8706
Ginie31	0,3049	0,5109	0,3049	1,2272
Greensky	0,3741	200,6167	0,3741	0,1376
gzing	0,2867	15,3597	0,2867	0,4569
hutch	0,3248	176,0672	0,3248	0,1865
jeanbask	0,1789	14439,6575	0,1789	0,1803
jeannot3	0,5694	53,2498	0,5694	0,1040
joewj	0,2461	6,0559	0,2461	0,6420
juboubou	2,0915	298,7582	2,0915	0,0000

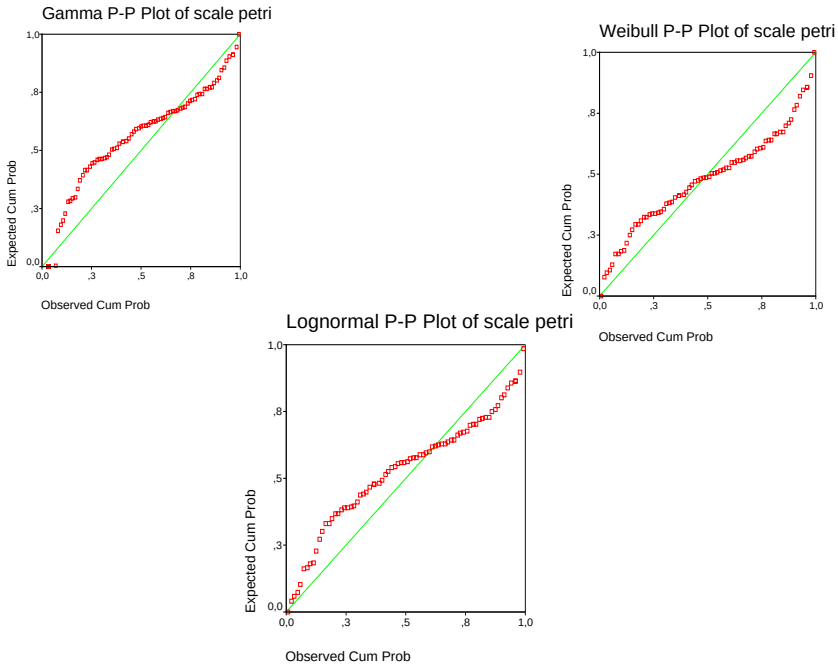
juju75	0,1000	4567,0000	0,1000	0,4306
kabotine	0,3906	17,1297	0,3906	0,3297
Kifran	0,2620	30,8200	0,2620	0,4073
lbarbier	0,4700	18,3040	0,4700	0,2551
levieux	0,3516	72,9181	0,3516	0,2213
Loquet	0,3198	38,9303	0,3198	0,3100
luberdov	0,2000	678,0000	0,2000	0,2715
m@hy	0,3635	23,9590	0,3635	0,3152
manwe	2,2227	519,6988	2,2227	0,0000
mary_ce	1,5539	343,2429	1,5539	0,0001
Mayfair	0,3507	55,0011	0,3507	0,2452
mekiela	0,4386	44,6049	0,4386	0,1890
Michel	0,1409	146,3940	0,1409	0,4954
minious	0,4225	37,4245	0,4225	0,2165
mperrot	0,3758	17,6528	0,3758	0,3399
Munier	0,3209	17,4933	0,3209	0,3992
Naemesis	0,2158	78,6591	0,2158	0,3899
nhai07	0,1000	1,0000	0,1000	1,0000
philia	0,2190	17,5582	0,2190	0,5340
popcorn	0,3799	27,8814	0,3799	0,2825
prioton	0,3706	935,9980	0,3706	0,0792
prozaiik	0,4103	10,1312	0,4103	0,3867
quentin	0,3435	34,7526	0,3435	0,2956
RAPH	0,1000	1,0000	0,1000	1,0000
rlaval	0,3433	61,7707	0,3433	0,2427
sandor	0,0100	92,0000	0,0100	0,9558
semery	0,3310	6,1453	0,3310	0,5483
sophynet	0,1987	22,9838	0,1987	0,5364
Stèph	0,2793	17,1902	0,2793	0,4518
stumpf	3,0000	666,0000	3,0000	0,0000
Sylvie4	8,0000	213,0000	8,0000	0,0000
Thibaut	0,4902	23,1901	0,4902	0,2142
toto2	0,3151	748,0455	0,3151	0,1243
tranb	0,3613	128,8617	0,3613	0,1728
vianney2	0,3404	30,7697	0,3404	0,3115
Virginie	0,3119	224,5162	0,3119	0,1848
xzouf	0,3083	5,5121	0,3083	0,5908
Yseulys	0,3916	12,2834	0,3916	0,3745
Yves	0,3782	122,6043	0,3782	0,1622

Nous constatons des paramètres qui diffèrent selon les individus, montrant ainsi une hétérogénéité du comportement de connexion des agents B.

Par conséquent, nous avons étudié la distribution de chacun des paramètres de la loi de Weibull représentant la durée entre deux connexions d'un agent B.

Plusieurs lois ont été testées, les plus proches des données observées sont présentées ci-dessous :

Figure 4.4 : PP Plots paramètre d'échelle alpha de la loi de Weibull



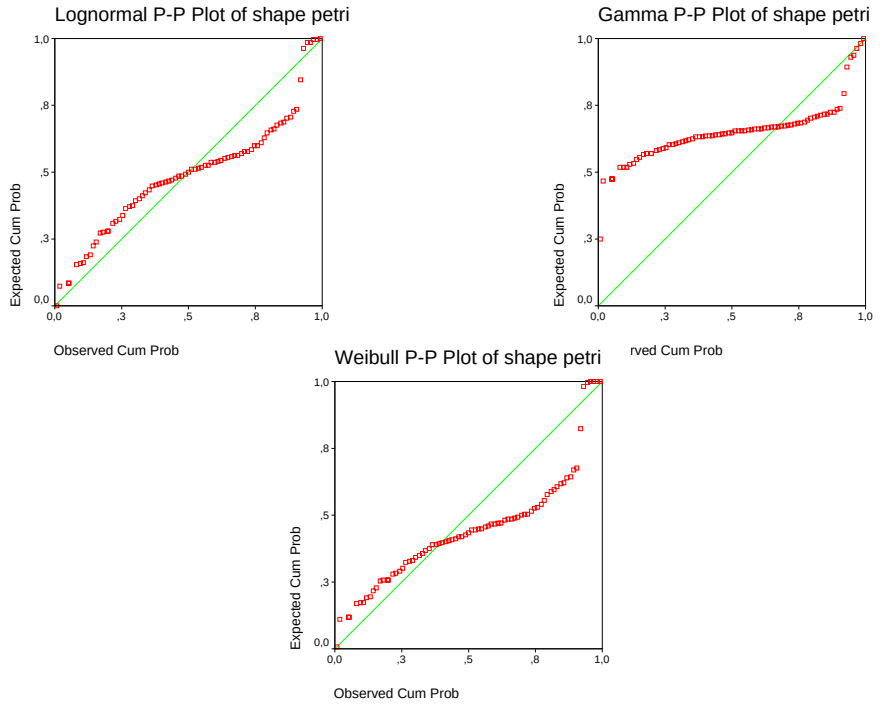
Ces courbes montrent des écarts par rapport aux données observées. La courbe la plus proche des données observées semble cependant être la loi de Weibull. Les lois Gamma et Weibull ont des vraisemblances très proches (– 205,32 pour Gamma vs -205,86 pour Weibull).
Les paramètres des lois ont été calculés sous excel en maximisant la vraisemblance et en utilisant la fonction solveur.

Afin de valider que le paramètre alpha suit une loi de Weibull et étant donné quelques écarts observés sur le PP plot de la figure 4.4, nous avons mené un test de Kolmogorov-Smirnov afin de valider que ce paramètre alpha suit bien une loi de Weibull.
Les résultats, présentés ci-dessous, montrent un résultat non significatif (Asymp.Sig.>0.05) nous permettant de ne pas rejeter l'hypothèse nulle et de conclure que les deux distributions (données observées d'alpha et loi de Weibull) ne sont pas différentes.

Two sample Kolmogorov-Smirnov Test

		shape petri
Most Extreme Differences	Absolute	,166
	Positive	,166
	Negative	-,097
Kolmogorov-Smirnov Z		1,038
Asymp. Sig. (2-tailed)		,232

Figure 4.5 : PP Plots paramètre de forme lambda de la loi de Weibull



En ce qui concerne le paramètre de forme lambda de la loi de Weibull représentant la durée entre deux connexions des agents B, on constate qu'une loi Gamma ne représente pas correctement les données observées (Figure 4.5). Les lois les moins mauvaises sont Weibull et Log-Normale dont les vraisemblances sont très proches.

Pour des raisons de cohérence de lois par rapport aux lois sélectionnées précédemment pour les agents A, nous avons testé la loi de Weibull. Là encore, nous avons effectué un test de

Kolmogorov-Smirnov afin de valider que cette distribution représente bien celle du paramètre lambda. Les résultats du test sont les suivants :

Two Sample Kolmogorov Test

		scale petri
Most Extreme Differences	Absolute	,178
	Positive	,177
	Negative	-,178
Kolmogorov-Smirnov Z		1,114
Asymp. Sig. (2-tailed)		,167

Les résultats, présentés ci-dessous, montrent un résultat non significatif (Asymp.Sig.>0.05) nous permettant de ne pas rejeter l'hypothèse nulle et de conclure que les deux distributions (données observées de lambda et loi de Weibull) ne sont pas différentes.

La durée entre deux connexions des agents B sur le site Millemercis distribution suit par conséquent une loi de Weibull de paramètres alpha et lambda, avec :

- alpha suit une loi Weibull de paramètres (formules adaptées au réseau de Petri) 1.4944 et 0.4167
- lambda suit une loi de Weibull de paramètre (formules adaptées au réseau de Petri) alpha 1 = 3.7629 et béta1=1.4378

Les paramètres ci-dessus ont été calculés sous excel en maximisant la vraisemblance et en utilisant la fonction solveur.

Nous avons également testé si les deux paramètres alpha et lambda était corrélés. Pour cela, nous avons effectué une régression linéaire du paramètre alpha sur le paramètre lambda. Les résultats sont fournis ci-dessous :

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	,760	,145		5,240	,000
	scale petri	-,565	,249	-,248	-2,275	,026

a. Dependent Variable: shape petri

On constate un lien significatif (Sig. = 0.026 < 0.05) mais en revanche, le R^2 global du modèle linéaire est de 0.061, ce qui montre que ce modèle ne représente pas correctement les données observées. On peut donc conclure que les deux paramètres ne sont que très peu corrélés.

Loi d'achat après une connexion :

Conformément à la méthode décrite dans le chapitre 3, nous avons ensuite estimé la loi d'achat des agents A et B après une connexion.
Les paramètres de la loi NBD permettant de représenter le comportement d'achat des agents A ou B après une connexion ont aussi été calculés sur une période de 12 mois d'historique (par rapport à la date de création de la première liste) et par la méthode de maximisation de la vraisemblance sur excel, pour chaque login étudié.
La distribution d'un modèle NBD est déterminé par la formule suivante (Hardie et Fader, 2001) :

$$P(X=x)=\frac{\Gamma(r+x)}{\Gamma(r)x!}\left(\frac{\alpha}{\alpha+1}\right)^r\left(\frac{1}{\alpha+1}\right)^x$$

α et r sont les paramètres de la loi NBD.
Les résultats des paramètres r et alpha sur chacun des agents sont fournis dans le tableau 4.3 ci-dessous :

Tableau 4.3: paramètres NBD		
	r	alpha
130961	18384,0039	2377532,4936
akwell1	0,0519	0,1719
AlexO	0,0677	0,9477
ama1	27140,4764	515631,9726
amibio	0,8000	34567789,0000
AnnA	0,0258	1,5024
Aude	0,0190	0,1386
badir	0,0000	19876542567,0000
bgougay	0,0830	2,6001
blondi	3971082,9506	93446861,0437
bpdl	0,0335	0,3274
bruyere	0,0001	123456789,0000
bschunke	0,1469	0,7641
carolala	0,0010	87676545678,0000
cathou1	23741871,1771	123457546,0930
cclerget	23741871,1771	123457546,0930
Céline21	215369,7273	3022810,1922
chap's	302114,2477	4223890,6454
Cypher	23741871,1771	123457546,0930
DALMATE	0,0010	123456789,0000
Delatte	7880264,3414	123456814,2828
didine77	0,0010	123456789,0000
Dinou	21470716,2143	123456874,1265
djoule	0,0115	0,5037
dolores3	5367708,4430	123456794,3201
doppel	0,2033	0,4067
eb95	14524383,5391	123456762,3852
eb96	9496682,8727	123456792,8654
Emilie	0,1000	987876765,0000
estia	0,0299	0,1893
eusebio1	0,0736	0,9814
fabou	0,1000	987876765,0000
fallouis	0,3035	2,6098
fevrard	4015180,6009	53687092,0382
FOURNIAL	0,2595	1,2543
Fruor	53687091,3000	987876764,9946
frzpower	0,0317	0,4273
GaëlleM	0,1981	0,8719
gallouin	0,0073	0,2647
Ginie31	0,0025	0,4003
Greensky	0,0917	0,6155
gzing	0,0086	76654543,0000
HERPIN	0,0000	76543218,0000
hutch	0,0176	0,4143
jeanbask	6709812,6582	76543245,0039
jeannot3	0,0192	0,7271

Tableau 4.3: paramètres NBD		
	r	alpha
jeda	0,0000	76543218,0000
JIMASA	0,0000	76543218,0000
joewj	0,0000	76543218,0000
juboubou	0,0000	76543218,0000
juju75	0,0000	76543218,0000
kabotine	0,0576	0,5448
Kifran	0,0000	987876765,0000
lbarbier	0,0000	1987627567,0000
levieux	0,0001	987654679,0000
Loquet	0,0000	987876765,0000
luberdov	0,0001	8767654,0000
m@hy	0,0000	876543567,0000
manwe	0,0000	87656743,0000
mary_ce	0,0001	9878765435,0000
Mayfair	0,0000	87676545678,0000
mekiela	0,0510	0,4333
Michel	0,0000	8767654567,0000
minious	0,0000	787657845,0000
mperrot	0,0266	0,9128
Munier	0,1000	987876765,0000
Naemesis	0,0000	987876578,0000
nhai07	0,0000	87654567389,0000
ninie12	0,1000	987876765,0000
philia	604555,6160	63465550,3311
picardjf	0,0000	53687092,0382
popcorn	0,0001	786765435,0000
prioton	0,0001	786765435,0000
prozaiik	0,0000	987876764,9900
quentin	0,0000	87676543567,0000
RAPH	0,0000	6787654678,0000
rlaval	2507,4898	876784,3837
romain37	0,0001	786756789,0000
sandor	0,0001	76654543,0000
semery	0,0113	0,8873
sophynet	0,0001	76543245,9896
Souris	0,0001	76543245,0039
starchen	0,0000	7656789,0000
Stèph	0,0000	76543218,0000
stumpf	0,0000	76543218,0000
Sylvie4	0,0000	76543218,0000
Thibaut	0,0051	0,4015
toto2	0,0000	76543218,0000
tranb	0,0001	56736789,0000
vianney2	0,0053	0,4010
Virginie	0,0001	767876546,0000
xzouf	619617,1770	111583352,4958
Yseulys	0,0085	0,2881
Yves	0,0001	987678765,0000

On constate sur le tableau 4.3 que les paramètres de la loi NBD sont très hétérogènes d'un login à l'autre. Ceci confirme les analyses que nous avons menées au chapitre 3 (partie 3.4.3) dans lesquelles nous avons étudié la répartition du nombre d'items achetés après une connexion. Nous avions montré que la majorité des logins comprend un nombre maximum d'items achetés supérieur à 10 mais que ce chiffre variait entre 10 et 64. L'hétérogénéité des paramètres de la loi NBD reflète le fait que sur certaines listes, il y a peu d'achats groupés alors que sur d'autres plusieurs items sont achetés en un coup après la connexion. D'autre part, il existe certaines listes avec très peu d'achats.

Nous avons donc tout d'abord essayé d'étudier l'hétérogénéité des paramètres de la loi NBD qui ressort dans le tableau 4.3. Pour cela, nous avons testé plusieurs lois possibles représentant la distribution du paramètre r et celle du paramètre alpha.

Les résultats des lois les plus proches des données observées sont présentés dans les figures ci-dessous.

Figure 4.6 : PP plots du paramètre r de la loi NBD

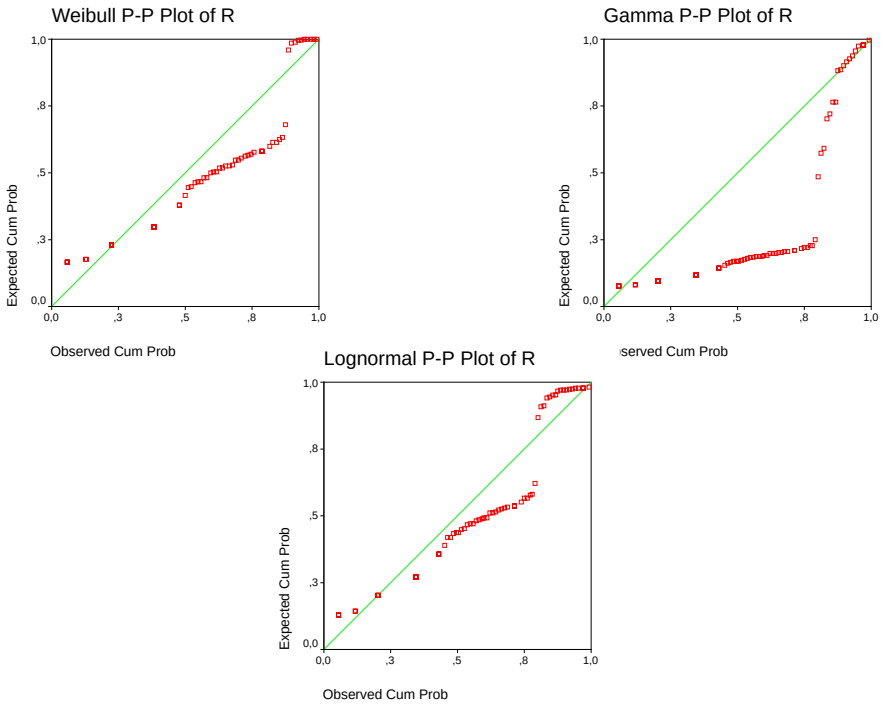
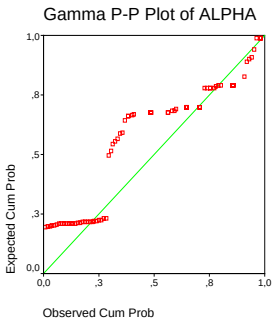
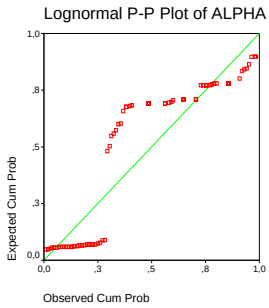
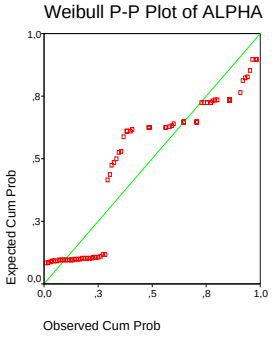


Figure 4.7 : PP plots du paramètre alpha de la loi NBD







On constate sur les deux paramètres r et alpha qu'aucune loi ne représente correctement les données observées, même si la loi de Weibull semble la plus proche.

Nous avons donc opté dans un deuxième temps pour une approche de segmentation en étudiant une loi NBD sur l'ensemble des agents dont les paramètres varient sur plusieurs segments d'individus, l'hétérogénéité étant dans ce cas représentée par les différents segments.

Nous avons testé plusieurs modèles possibles sur l'ensemble des individus pour la loi d'achat après connexion sur l'ensemble des individus :

- un modèle de Poisson
- un modèle NBD
- un modèle ZNBD (le principe de ce modèle a déjà été présenté dans le chapitre 3), en raison du nombre important de non achats)
- un modèle 2 segments NBD
- un modèle 3 segments NBD
- un modèle 4 segments NBD
- un modèle 5 segments NBD

Chacun de ces modèles a été estimé sous excel en maximisant la fonction de vraisemblance par la fonction solveur (Hardie et Fader, 2001).

Nous avons comparé ces modèles en utilisant le critère d'information bayésien (critère BIC, Hardie et Fader, 2001).

BIC = - 2LL + p ln(N)

avec :

p nombre de paramètres du modèle

N taille de l'échantillon

Le choix du meilleur modèle s'effectue en minimisant le BIC.

Les résultats sur les modèles testés sont les suivants :

Tableau 4.4 : comparaison des modèles NBD

Model	LL	# params	BIC	X^2 p-value
Poisson	-7280,3701	1	14565,0915	0,0000
NBD	-3501,1366	2	7010,9759	0,0000
ZNBD	-3500,1300	3	7013,3141	0,0000
2 seg NBD	-3478,1124	5	6977,9814	0,0000
3 seg NBD	-3476,8779	9	6992,7002	0,2846
4 seg NBD	-3476,4473	11	7000,7592	0,5744
5 seg NBD	-3476,7725	14	7014,4636	0,7824

Le tableau 4.4 montre que le meilleur modèle est le modèle NBD avec 3 segments. En effet, au delà de 3 segments, le BIC est supérieur à 6992 (BIC pour le modèle à 3 segments). Le modèle à 2 segments ne donne pas de test du X2 significatif.

Les formules et les résultats sous excel des paramètres du modèle NBD avec 3 segments sont :

Tableau 4.5 : Formules du modèle NBD 3 segments

r	0,35	s				**proba achat aprè**
alpha	1,63				LL	=SOMME(H14:H3
r	0,03	s				
alpha	0,16					
pi_1	0,15					
pi_2	0,56					
r	-0,0	s				
alpha	0,26					
pi_3	="1-					
BIC	="-2	=				
		P(X=x)				

Nbre it	Nbre	Seg1	Seg2	Seg3	Total	LL
0	219(	=(B2/(B2+1))^B1	=(B4/(B4+1))^B3	=(B8/(B8+1))^B7	=D14*B$5+B$6*E14+(1-B$5-B$6)*F14	=B14*LN(G14)
1	321	=(B$1+$A15-1)/($A15*(B$2+1))*D14	=(B$3+$A15-1)/($A15*(B$4+1))*E14	=(B$7+$A15-1)/($A15*(B$8+1))*F14	=D15*B$5+B$6*E15+(1-B$5-B$6)*F15	=B15*LN(G15)
2	96	=(B$1+$A16-1)/($A16*(B$2+1))*D15	=(B$3+$A16-1)/($A16*(B$4+1))*E15	=(B$7+$A16-1)/($A16*(B$8+1))*F15	=D16*B$5+B$6*E16+(1-B$5-B$6)*F16	=B16*LN(G16)
3	42	=(B$1+$A17-1)/($A17*(B$2+1))*D16	=(B$3+$A17-1)/($A17*(B$4+1))*E16	=(B$7+$A17-1)/($A17*(B$8+1))*F16	=D17*B$5+B$6*E17+(1-B$5-B$6)*F17	=B17*LN(G17)
4	19	=(B$1+$A18-1)/($A18*(B$2+1))*D17	=(B$3+$A18-1)/($A18*(B$4+1))*E17	=(B$7+$A18-1)/($A18*(B$8+1))*F17	=D18*B$5+B$6*E18+(1-B$5-B$6)*F18	=B18*LN(G18)
5	12	=(B$1+$A19-1)/($A19*(B$2+1))*D18	=(B$3+$A19-1)/($A19*(B$4+1))*E18	=(B$7+$A19-1)/($A19*(B$8+1))*F18	=D19*B$5+B$6*E19+(1-B$5-B$6)*F19	=B19*LN(G19)
6	15	=(B$1+$A20-1)/($A20*(B$2+1))*D19	=(B$3+$A20-1)/($A20*(B$4+1))*E19	=(B$7+$A20-1)/($A20*(B$8+1))*F19	=D20*B$5+B$6*E20+(1-B$5-B$6)*F20	=B20*LN(G20)
7	10	=(B$1+$A21-1)/($A21*(B$2+1))*D20	=(B$3+$A21-1)/($A21*(B$4+1))*E20	=(B$7+$A21-1)/($A21*(B$8+1))*F20	=D21*B$5+B$6*E21+(1-B$5-B$6)*F21	=B21*LN(G21)
8	2	=(B$1+$A22-1)/($A22*(B$2+1))*D21	=(B$3+$A22-1)/($A22*(B$4+1))*E21	=(B$7+$A22-1)/($A22*(B$8+1))*F21	=D22*B$5+B$6*E22+(1-B$5-B$6)*F22	=B22*LN(G22)
9	7	=(B$1+$A23-1)/($A23*(B$2+1))*D22	=(B$3+$A23-1)/($A23*(B$4+1))*E22	=(B$7+$A23-1)/($A23*(B$8+1))*F22	=D23*B$5+B$6*E23+(1-B$5-B$6)*F23	=B23*LN(G23)
10	6	=(B$1+$A24-1)/($A24*(B$2+1))*D23	=(B$3+$A24-1)/($A24*(B$4+1))*E23	=(B$7+$A24-1)/($A24*(B$8+1))*F23	=D24*B$5+B$6*E24+(1-B$5-B$6)*F24	=B24*LN(G24)
11	2	=(B$1+$A25-1)/($A25*(B$2+1))*D24	=(B$3+$A25-1)/($A25*(B$4+1))*E24	=(B$7+$A25-1)/($A25*(B$8+1))*F24	=D25*B$5+B$6*E25+(1-B$5-B$6)*F25	=B25*LN(G25)
12	3	=(B$1+$A26-1)/($A26*(B$2+1))*D25	=(B$3+$A26-1)/($A26*(B$4+1))*E25	=(B$7+$A26-1)/($A26*(B$8+1))*F25	=D26*B$5+B$6*E26+(1-B$5-B$6)*F26	=B26*LN(G26)
13	4	=(B$1+$A27-1)/($A27*(B$2+1))*D26	=(B$3+$A27-1)/($A27*(B$4+1))*E26	=(B$7+$A27-1)/($A27*(B$8+1))*F26	=D27*B$5+B$6*E27+(1-B$5-B$6)*F27	=B27*LN(G27)
14	4	=(B$1+$A28-1)/($A28*(B$2+1))*D27	=(B$3+$A28-1)/($A28*(B$4+1))*E27	=(B$7+$A28-1)/($A28*(B$8+1))*F27	=D28*B$5+B$6*E28+(1-B$5-B$6)*F28	=B28*LN(G28)
15	1	=(B$1+$A29-1)/($A29*(B$2+1))*D28	=(B$3+$A29-1)/($A29*(B$4+1))*E28	=(B$7+$A29-1)/($A29*(B$8+1))*F28	=D29*B$5+B$6*E29+(1-B$5-B$6)*F29	=B29*LN(G29)
16	1	=(B$1+$A30-1)/($A30*(B$2+1))*D29	=(B$3+$A30-1)/($A30*(B$4+1))*E29	=(B$7+$A30-1)/($A30*(B$8+1))*F29	=D30*B$5+B$6*E30+(1-B$5-B$6)*F30	=B30*LN(G30)
17	1	=(B$1+$A31-1)/($A31*(B$2+1))*D30	=(B$3+$A31-1)/($A31*(B$4+1))*E30	=(B$7+$A31-1)/($A31*(B$8+1))*F30	=D31*B$5+B$6*E31+(1-B$5-B$6)*F31	=B31*LN(G31)
20	1	=(B$1+$A32-1)/($A32*(B$2+1))*D31	=(B$3+$A32-1)/($A32*(B$4+1))*E31	=(B$7+$A32-1)/($A32*(B$8+1))*F31	=D32*B$5+B$6*E32+(1-B$5-B$6)*F32	=B32*LN(G32)
21	2	=(B$1+$A33-1)/($A33*(B$2+1))*D32	=(B$3+$A33-1)/($A33*(B$4+1))*E32	=(B$7+$A33-1)/($A33*(B$8+1))*F32	=D33*B$5+B$6*E33+(1-B$5-B$6)*F33	=B33*LN(G33)
22	2	=(B$1+$A34-1)/($A34*(B$2+1))*D33	=(B$3+$A34-1)/($A34*(B$4+1))*E33	=(B$7+$A34-1)/($A34*(B$8+1))*F33	=D34*B$5+B$6*E34+(1-B$5-B$6)*F34	=B34*LN(G34)
24	1	=(B$1+$A35-1)/($A35*(B$2+1))*D34	=(B$3+$A35-1)/($A35*(B$4+1))*E34	=(B$7+$A35-1)/($A35*(B$8+1))*F34	=D35*B$5+B$6*E35+(1-B$5-B$6)*F35	=B35*LN(G35)
25	1	=(B$1+$A36-1)/($A36*(B$2+1))*D35	=(B$3+$A36-1)/($A36*(B$4+1))*E35	=(B$7+$A36-1)/($A36*(B$8+1))*F35	=D36*B$5+B$6*E36+(1-B$5-B$6)*F36	=B36*LN(G36)
26	1	=(B$1+$A37-1)/($A37*(B$2+1))*D36	=(B$3+$A37-1)/($A37*(B$4+1))*E36	=(B$7+$A37-1)/($A37*(B$8+1))*F36	=D37*B$5+B$6*E37+(1-B$5-B$6)*F37	=B37*LN(G37)
64	1	=(B$1+$A38-1)/($A38*(B$2+1))*D37	=(B$3+$A38-1)/($A38*(B$4+1))*E37	=(B$7+$A38-1)/($A38*(B$8+1))*F37	=D38*B$5+B$6*E38+(1-B$5-B$6)*F38	=B38*LN(G38)

Tableau 4.6 : Résultats du modèle NBD 3 segments

r	0,1765	segment 1
alpha	1,1814	
r	2,1333	segment 2
alpha	0,2767	
pi_1	0,1760	
pi_2	0,0042	
r	0,0019	segment 3
alpha	0,2094	
pi_3	0,8198	

3 Segments

proba achat après une connexion

LL -3476,8779

BIC -2LL+plogN 6992,917739

P(X=x)

Nbre iter	Nbre cas	Seg1	Seg2	Seg3	Total	LL	Expected
0	21901	0,8974	0,0383	0,9968	0,9753	-548,1000	21900,9844
1	321	0,0726	0,0640	0,0015	0,0143	-1363,5167	321,0502
2	96	0,0196	0,0785	0,0006	0,0043	-523,3132	96,3641
3	42	0,0065	0,0848	0,0003	0,0018	-265,7935	40,0807
4	19	0,0024	0,0852	0,0002	0,0009	-132,2359	21,3176
5	12	0,0009	0,0819	0,0001	0,0006	-88,6697	13,8761
6	15	0,0004	0,0762	0,0001	0,0005	-115,2179	10,3617
7	10	0,0001	0,0694	0,0001	0,0004	-78,9741	8,3470
8	2	0,0001	0,0620	0,0001	0,0003	-16,1562	6,9672
9	7	0,0000	0,0547	0,0000	0,0003	-57,7132	5,8977
10	6	0,0000	0,0477	0,0000	0,0002	-50,4459	5,0112
11	2	0,0000	0,0412	0,0000	0,0002	-17,1428	4,2541
12	3	0,0000	0,0353	0,0000	0,0002	-26,2141	3,6013
13	4	0,0000	0,0301	0,0000	0,0001	-35,6325	3,0379
14	4	0,0000	0,0255	0,0000	0,0001	-36,3276	2,5534
15	1	0,0000	0,0215	0,0000	0,0001	-9,2592	2,1385
16	1	0,0000	0,0180	0,0000	0,0001	-9,4398	1,7851
17	1	0,0000	0,0150	0,0000	0,0001	-9,6235	1,4855
20	1	0,0000	0,0125	0,0000	0,0001	-9,8154	1,2261
21	2	0,0000	0,0103	0,0000	0,0000	-20,0190	1,0099
22	2	0,0000	0,0085	0,0000	0,0000	-20,4111	0,8301
24	1	0,0000	0,0069	0,0000	0,0000	-10,4053	0,6798
25	1	0,0000	0,0057	0,0000	0,0000	-10,6066	0,5559
26	1	0,0000	0,0046	0,0000	0,0000	-10,8093	0,4538
64	1	0,0000	0,0037	0,0000	0,0000	-11,0351	0,3621

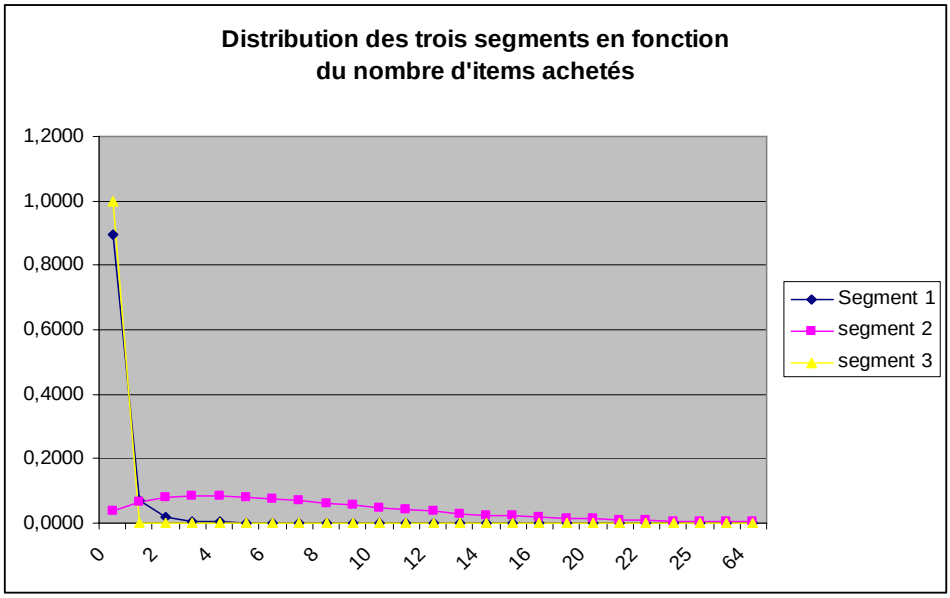
Nous pouvons donc conclure que sur l'ensemble des individus, le nombre d'items achetés après une connexion suit une loi NBD avec 3 segments, de paramètres :

Tableau 4.7 : Paramètres du modèle NBD avec 3 segments

r	0,1765	segment 1
alpha	1,1814	
r	2,1333	segment 2
alpha	0,2767	
pi_1	0,1760	
pi_2	0,0042	
r	0,0019	segment 3
alpha	0,2094	
pi_3	0,8198	

Le tableau 4.7 montre que le segment 1 représente 17.6% des individus, le segment 2 correspond à 0.4% des individus tandis que le segment 3 couvre 82% des individus. Ces trois groupes ont des comportements d'achat après connexion différents, cette hétérogénéité étant captée par les paramètres r et alpha spécifiques à chaque segment et fournis dans le tableau 4.7. Le segment 2 représentant une part très faible des agents, nous aurions pu nous limiter à deux segments. Cependant, le segment 2 a des caractéristiques spécifiques intéressantes qui justifient sont analyse.

La distribution de la loi NBD sur ces trois segments est présentée sur la figure ci-dessous.



On peut interpréter les trois groupes de la manière suivante : le segment 1 est essentiellement constitué d'agents qui n'achètent pas d'item (89% des cas) ou un seul item (7 ,7% des cas). Le segment 2 en revanche est constitué d'agents qui achètent plusieurs items à la fois (mais ce groupe représente une part très faible des agents). Enfin, le segment 3 est majoritairement constitué d'agents qui sont non acheteurs (99.7% des cas).

Nous avons donc deux possibilités de modélisation sur le réseau de Petri :

- Une approche par individu, dans laquelle chaque agent A est caractérisé par ses propres paramètres (2 paramètres pour la loi de Weibull représentant la durée interconnexion agent A, 2 paramètres pour la loi de Weibull représentant la durée interconnexion agent B, 2 paramètres pour la loi NBD d'achat d'items sur la liste).
- Une approche agrégée, dans laquelle on distingue 3 segments d'agents A.
    - Chaque segment a ses propres paramètres pour la loi NBD (tableau 4.7),
    - La durée entre deux connexions des agents A suit une loi de Weibull de paramètre alpha et lambda, alpha suivant une loi Gamma (paramètres 0.8265 et 3.9766) et lambda une loi de Weibull (paramètres 3.6099 et 1.6180)
    - La durée entre deux connexions des agents B suit une loi de Weibull de paramètre alpha et lambda, alpha suivant une loi Gamma (paramètres 1.9618 et 5.4169) et lambda une loi de Weibull (paramètres 3.7629 et 1.4378)

Le package SPNP (développé par Duke University) que nous utilisons pour les réseaux de Petri ne permettant pas actuellement de programmer les fonctions Weibull et Gamma pour la distribution des paramètres de la loi de Weibull de connexion des agents A et B, nous avons choisi la première approche qui est une modélisation individuelle pour chacun des 94 agents. Les deux principaux avantages de ce choix sont tout d'abord le fait d'avoir une représentation plus précise du phénomène étudié et d'autre part de refléter l'hétérogénéité des individus qui est importante, comme nous l'avons souligné précédemment. La contrepartie est le nombre important de paramètres à estimer puisque 6 paramètres doivent être estimés par agent avec parfois peu de données (loi de Weibull durée inter connexion des agents A, loi de Weibull durée inter connexion des agents B, loi NBD pour le nombre d'items achetés).

Pour notre modèle par individu, le paramétrage de l'achat dans le modèle du réseau de Petri a été effectué en utilisant les probabilités d'achat (avec nbre d'items achetés) ou de non achat

Tableau 4.8: loi d'achat

	Nbre items achetes après connexion								Nbre total connexions
	0	1	2	3	4	5	6	8	
130961	316	3							319
akwell1	43	2	1				2		48
AlexO	34	2		1					37
ama1	10								10
amibio	24								24
AnnA	809	9	2		1				821
Aude	34						1		35
badir	18								18
bgougay	163	4	1						168
blondi	17								17
bpdl	66	1	2		1				70
bruyere	23								23
bschunke	20	1		1					22
carolala	41								41
cathou1	3								3
cclerget	3	1							4
Céline21	6	2							8
chap's	16	3							19
Cypher	21	1			1				23
DALMATE	6								6
Delatte	35	2							37
didine77	24								24
Dinou	17	4							21
djoule	896	10	1					1	908
dolores3	19								19
doppel	11	2	1			1			15
eb95	25	1	1						27
eb96	9	1							10
Emilie	825								825
estia	16								16
eusebio1	26								26
fabou	14								14
fallouis	38	3	1						42
fevrard	16	1							17
FOURNIAL	19	1	2						22
Fruor	12	1							13
frzpower	23								23
GaëlleM	9	1	1						11
gallouin	538	2	1						541
Ginie31	6								6
Greensky	32	3			1				36
gzing	69								69
HERPIN	28								28
hutch	40								40
jeanbask	5								5

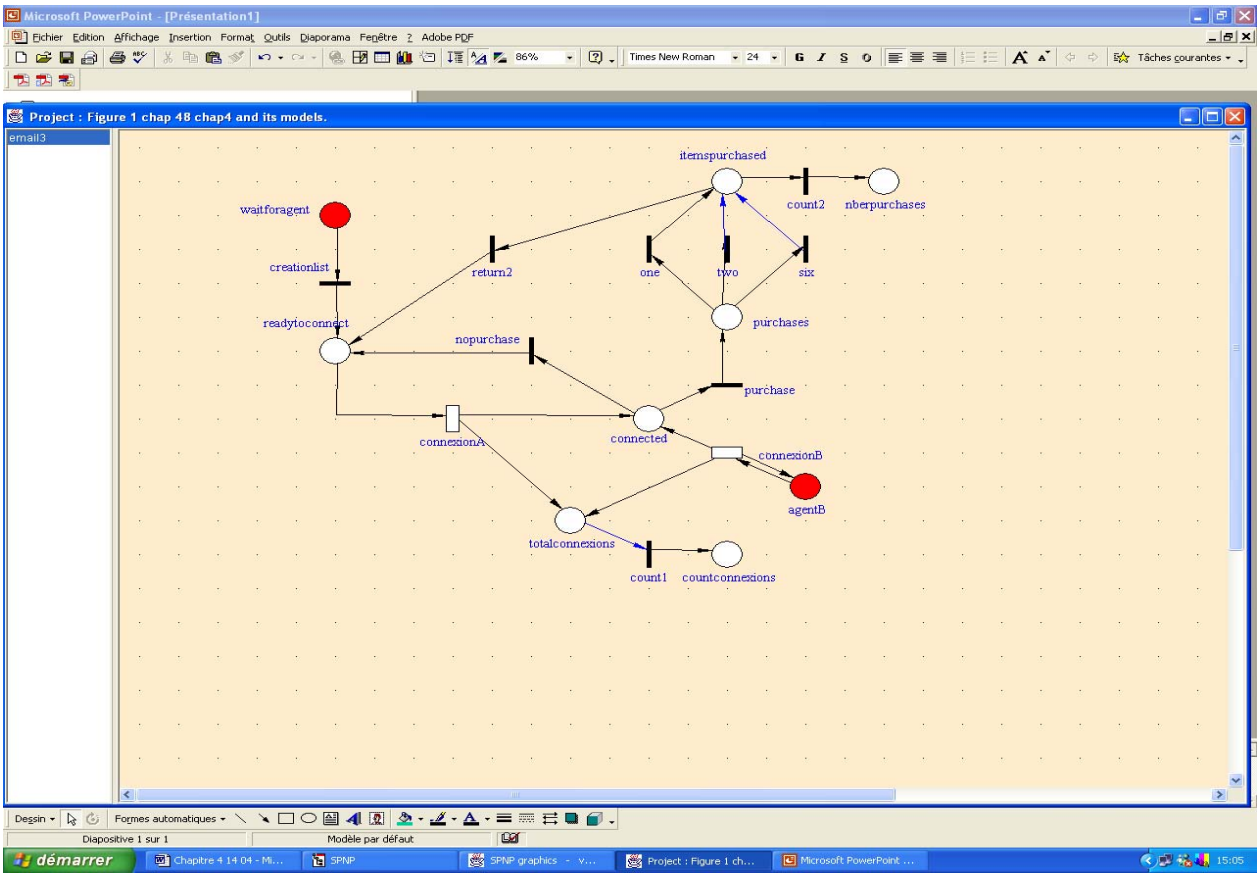
pour les transitions *purchase/no purchase,* en utilisant les tableaux ci-dessous par individu.

Tableau 4.8: loi d'achat

	Nbre items achetes après connexion								Nbre total connexions
	0	1	2	3	4	5	6	8	
jeannot3	730	7	1				1		739
jeda	16								16
JIMASA	8								8
joewj	15								15
juboubou	16								16
juju75	36								36
kabotine	43								43
Kifran	43								43
lbarbier	78								78
levieux	16								16
Loquet	12								12
luberdov	4								4
m@hy	89								89
manwe	12								12
mary_ce	26								26
Mayfair	18								18
mekiela	32	1		1					34
Michel	20								20
minious	39								39
mperrot	62	1	1						64
Munier	22								22
Naemesis	33								33
nhai07	88								88
ninie12	14								14
philia	50	1							51
picardjf	5								5
popcorn	72								72
prioton	18								18
prozaiik	11								11
quentin	14								14
RAPH	27								27
rlaval	972	2							974
romain37	17								17
sandor	9								9
semery	708	1							709
sophynet	13								13
Souris	33								33
starchen	5								5
Stèph	33								33
stumpf	24								24
Sylvie4	17								17
Thibaut	495	1		2					498
toto2	6								6
tranb	15								15
vianney2	440	2			1				443
Virginie	47								47
xzouf	4								4
Yseulys	289	2	1						292
Yves	18								18

Considérons l'exemple du login akwell1 de la figure 1

Figure 4.8: réseau de Petri du login akwell1



Le principe de ce réseau de Petri a déjà présenté au chapitre précédent mais re précisons ici ses grandes lignes. Après avoir créé sa liste (transition *creationlist*), l'agent A se connecte sur sa wish list (transition *connexionA*) avec une durée entre deux connexion qui suit une loi de Weibull. D'autre part, les agents B peuvent également se connecter sur la wish list de l'agent A (transition *connexionB*) avec une durée entre deux connexions des agents B qui suit une loi de Weibull également. Après l'une ou l'autre de ces connexions, l'agent (A ou B) achète (transition *purchase*) ou n'achète pas (transition *nopurchase*). Dans le cas d'un achat, on calcule le nombre d'items achetés (transitions *one two* et *six* pour le login akwell1) qui constituera la place *itemspurchased*. Les jetons de cette place sont ensuite dirigés dans la place *nbrepurchases* qui permet de mesurer de manière cumulée le nombre d'items achetés .

En parallèle, un jeton retourne dans la place *readytoconnect* signifiant que l'agent A après son achat quitte le site millemercis et qu'il est à nouveau dans l'état où il peut se reconnecter sur la liste de cadeaux.

4.2.2 Performance du modèle réseau de Petri:

Afin d'étudier l'adéquation du modèle de Petri par rapport aux données observées sur la variable modélisée (nombre d'items achetés), nous avons utilisé les deux indicateurs classiques suivant:

MAD (Mean Absolute Deviation) : Somme des différences entre les données observées et les valeurs estimées par le modèle, sur l'ensemble des individus.

MSE (Mean Squared Errors): Somme du carré des différences entre les données observées et les valeurs estimées par le modèle, divisée par le nombre d'individus.

La période sur laquelle a été calculée ces indicateurs est la période de calibration de 12 mois.

Les résultats sont fournis par login dans le tableau 4.9 suivant :

Tableau 4.9: Performance des réseaux de Pétri

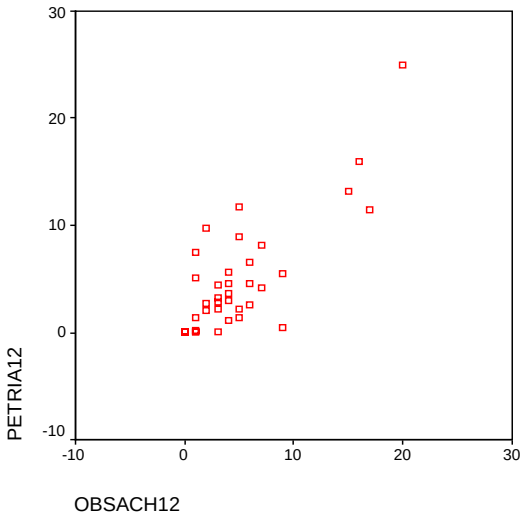
Login	Nbre d'items achetés Observés	Achats Petri	Observ-Modèle	(Observ-Modèle)^2
130961	3	2,19	0,81	0,66
akwell1	16	16,02	-0,02	0
AlexO	5	11,78	-6,78	46
ama1	0	0	0	0
amibio	0	0	0	0
AnnA	17	11,49	5,52	30,42
Aude	6	6,57	-0,57	0,32
badir	0	0	0	0
bgougay	6	2,58	3,42	11,7
blondi	0	0	0	0
bpdl	9	5,51	3,49	12,17
bruyere	0	0	0	0
bschunke	4	1,13	2,87	8,24
carolala	0	0	0	0
cathou1	0	0	0	0
cclerget	1	0,1	0,9	0,8
Céline21	2	9,68	-7,68	59,04
chap's	3	3,25	-0,25	0,06
Cypher	5	2,17	2,83	8,03
DALMATE	0	0	0	0
Delatte	2	2,68	-0,67	0,46
didine77	0	0	0	0
Dinou	4	4,62	-0,62	0,38
djoule	20	24,9	-4,9	24,04
dolores3	0	0	0	0
doppel	9	0,45	8,55	73,14
eb95	3	0	3	9
eb96	1	0,04	0,96	0,93
Emilie	0	0	0	0
estia	0	0	0	0
eusebio1	0	0	0	0
fabou	0	0	0	0
fallouis	5	1,4	3,6	12,94
fevrard	1	0,21	0,79	0,62
FOURNIAL	5	9	-4	16,02
Fruor	1	7,5	-6,5	42,25
frzpower	0	0	0	0
GaëlleM	3	2,8	0,2	0,04
gallouin	4	3,04	0,96	0,92
Ginie31	0	0	0	0
Greensky	7	4,18	2,82	7,97
gzing	0	0	0	0
HERPIN	0	0	0	0
hutch	0	0	0	0
jeanbask	0	0	0	0
jeannot3	15	13,18	1,82	3,32

Tableau 4.9: Performance des réseaux de Pétri calibration 12mois

Login	Nbre d'items achetés Observés	Achats Petri	Observ-Modèle	(Observ-Modèle)^2
jeda	0	0	0	0
JIMASA	0	0	0	0
joewj	0	0	0	0
juboubou	0	0	0	0
juju75	0	0	0	0
kabotine	0	0	0	0
Kifran	0	0	0	0
lbarbier	0	0	0	0
levieux	0	0	0	0
Loquet	0	0	0	0
luberdov	0	0	0	0
m@hy	0	0	0	0
manwe	0	0	0	0
mary_ce	0	0	0	0
Mayfair	0	0	0	0
mekiela	4	5,58	-1,58	2,49
Michel	0	0	0	0
minious	0	0	0	0
mperrot	3	4,39	-1,39	1,94
Munier	0	0	0	0
Naemesis	0	0	0	0
nhai07	0	0	0	0
ninie12	0	0	0	0
philia	1	1,45	-0,45	0,2
picardjf	0	0	0	0
popcorn	0	0	0	0
prioton	0	0	0	0
prozaiik	0	0	0	0
quentin	0	0	0	0
RAPH	0	0	0	0
rlaval	2	2,1	-0,1	0,01
romain37	0	0	0	0
sandor	0	0	0	0
semery	1	5,08	-4,08	16,64
sophynet	0	0	0	0
Souris	0	0	0	0
starchen	0	0	0	0
Stèph	0	0	0	0
stumpf	0	0	0	0
Sylvie4	0	0	0	0
Thibaut	7	8,14	-1,13	1,29
toto2	0	0	0	0
tranb	0	0	0	0
vianney2	6	4,58	1,42	2,02
Virginie	0	0	0	0
xzouf	0	0	0	0
Yseulys	4	3,65	0,35	0,13
Yves	0	0	0	0

On peut synthétiser ce tableau par la figure ci-dessous, présentant en ordonnées les valeurs estimées par le modèle de Pétri en fonction en abscisse des données observées d'achat (sur la période de calibration de 12 mois) pour chaque observation (une observation correspondant à un login)

Estimation du modèle de Pétri en fonction des données observées
(période de calibration de 12 mois)



Le tableau 4.9 montre globalement peu d'écart entre les données observées le modèle de Petri proposé. Les plus gros écarts concernent trois login et s'explique par le faible nombre d'observations (à savoir le nombre de connexions) dans chaque cas: Doppel (15 connexions) ; Céline21 (8 connexions) et Fruor (13 connexions).
Le calcul des 2 indicateurs MAD et MSE donne le résultat suivant :

Petri	MAD	4,19
12 mois	MSE	0,04

Tableau 4.10: performance du modèle

Le calcul du R^2 donne un résultat de 0.75, ce qui confirme que les réseaux de Petri représentent correctement les données observées sur lesquelles les paramètres avaient été évalués au niveau individuel. Rappelons ici que ces résultats portent sur les données qui ont permis l'estimation.

4.2.3 Comparaison du modèle du réseau de Petri par rapport à un modèle NBD:

L'objectif de cette partie est de déterminer si les réseaux de Petri constituent une méthode de modélisation plus pertinente que des modèles plus classiquement utilisés en Marketing.
L'une des variables modélisées dans le cadre des wish lists étant le nombre d'items achetés après une connexion, nous avons cherché à modéliser cette variable par un modèle classique NBD (Negative Binomial Distribution) avec régression sur des variables explicatives.
En effet, l'intérêt d'un modèle NBD « classique », avec régression, est de déterminer si certains facteurs expliquent le nombre d'items achetés. C'est un modèle plus explicatif par rapport au réseau de Petri dont l'objectif est de décrire avec précision les évènements séquentiels survenant dans un système, sans variable explicative extérieure (comme les variables socio-démographiques par exemple que nous n'avons pas intégrées dans le modèle de Petri.)

Rappelons que le modèle NBD avec régression se définit par la distribution suivante :

$$P(Yi=y)=\frac{\Gamma(r+y)}{\Gamma(r)y!}\left(\frac{\alpha}{\alpha+\exp(\beta' x_i}\right)^r\left(\frac{\exp(\beta' x_i)}{\alpha+\exp(\beta' x_i)}\right)^x$$

α et r sont les paramètres de la loi NBD et les variables x_i sont les variables explicatives utilisées dans la régression.

Comparaison des modèles sur la période de calibration

Nous avons par conséquent, par la méthode de maximisation de la vraisemblance sous excel, calculé le nombre d'items achetés en appliquant le modèle NBD avec régression. Les calculs sur excel (sur une période de 12 mois) nous permettent d'obtenir les 2 paramètres du modèle NBD par login permettant de faire des prévisions (12-16mois).
Les variables que nous avons intégrées dans le modèle NBD avec régression sont le nombre de listes par agent, le nombre de connexions des agent A et B, le profil socio-démographique (âge et sexe). Le choix de ces variables a été effectué à partir de l'hypothèse selon laquelle ces variables sont susceptibles de prédire le comportement d'achat des internautes sur les wish

lists. Le fichier excel permettant de calculer les résultats du modèle NBD est fourni ci-dessous. A gauche du fichier se trouvent les variables explicatives utilisées dans le modèle pour chaque login (Id dans le fichier excel). Ces variables sont connexions (nombre de connexions), sexe, nbre listes (nombre de listes) et Age. Nous avons pris la fonction log et ajouté 1 à la variable connexion (avant d'en prendre le log) afin d'éviter les valeurs des cases où il n'y a aucune connexion (ces cases seraient ressorties comme erreurs dans le modèle sinon). Nous avons également pris le log de la variable âge. A droite du fichier excel, sont indiqués les pourcentages trouvés pour le nombre d'items achetés par login (0,1, 2,…, 10+).

Les 3 pages suivantes correspondent au fichier chap4page1921.doc ci-joint.

Nous avons évalué le modèle NBD avec régression grâce au logiciel Stata. Ceci nous permet d'obtenir les erreurs (standard errors) et les p-value des paramètres du modèle NBD.

Les résultats sont fournis dans le tableau 4.11 ci-dessous.

Tableau 4.11: Estimation du modèle NBD avec régression

NBD regression Stata

LL=-174,0415

	Coef	S.E.	p-value
Ln Connexion	0,5648	0,1612	0,000
Ln Age	-0,4929	0,1673	0,003
Sexe	0,1349	0,3658	0,712
Nbre Listes	0,0752	0,1282	0,557
Dispersion	2,3402		

L'indicateur de dispersion dans l'analyse Stata montre le degré de dispersion des données. « d » équivaut en fait à $1+d*\exp(X_i\beta')$. Si d est proche de zéro, le modèle NBD correspond à une distribution de Poisson. Comme d>1, cela signifie que les données nombre d'items achetés sont très dispersées.

Le tableau 4.11 montre que seules deux variables sont significatives : le nombre de connexion et l'âge (p-value > 0,05 pour les variables Gender et Lists). Ceci signifie que l'âge et le nombre de connexions n'influent pas sur le nombre d'items achetés dans le cadre des wish lists. Comme les variables Sexe (Gender) et Age ne sont pas significatives, nous les avons exclues dans le modèle de prédiction.

Les paramètres du modèle NBD estimés sont les suivants :

Tableau 4.12: Paramètres de la loi NBD avec régression

alpha	0,9121
r	0,4234

Notre objectif étant de comparer les performances des modèles NBD et Petri, nous avons comparé les résultats des deux modèles sur les critères MAD et MSE que nous avons présentés précédemment. Nous avons calculé ces critères sur la variable nombre total des achats et sur la période de calibration de 12 mois, à partir des données qui nous ont servi à calibrer le modèle. Les résultats sont les suivants :

Tableau 4.13 : Comparaison NBD/ PETRI sur la période de calibration 12 mois

Nombre total des achats

MSE

12 months	NBD regression	12,97
	PETRI	4,19

MAD

12 months	NBD regression	1,15
	PETRI	0,04

On constate dans le tableau 4.13 que sur la période de calibration du modèle (les 12 premiers mois suivant la date de création de la première liste), les réseaux de Petri sont plus performants que le modèle NBD avec régression sur la variable nombre total des achats. En effet, les écarts entre les données observées sur la variable nombre total des achats et le modèle (reflétés par les indicateurs MAD et MSE) sont plus faibles pour Petri que pour NBD, en particulier pour l'indicateur MAD (0,04 vs 1,15). On constate également un écart important entre les critères MSE et MAD qui s'explique par un cas aberrant (le login Fruor).

Les meilleures performances du modèle Petri versus le modèle NBD avec régression, sur la variable nombre total des achats, peuvent s'expliquer par le fait que les réseaux de Petri décomposent un processus en évènements séquentiels et tiennent par conséquent compte de plus d'éléments que le modèle NBD qui tient surtout compte de variables explicatives de profils ou du comportement de connexions dans sa globalité mais pas séquentiellement. Plusieurs arguments peuvent être soulignés.

Tout d'abord, le modèle NBD utilise un nombre total de connexions alors que le réseau de Petri calcule après chaque connexion la probabilité d'acheter. D'autre part, chaque transition dans le modèle de Petri étant définie par la meilleure loi stochastique possible, les réseaux de Petri présentent l'avantage d'être particulièrement flexibles pour la calibration du modèle. Si les lois des transitions sont bien choisies, le modèle obtient de bons résultats. Cependant, deux limites méritent d'être soulignées. Tout d'abord nous comparons les modèles sur une seule variable (nombre total des achats). D'autre part, rappelons que le nombre de paramètres dans le modèle de Pétri est élevé (6 paramètres par individu).

Comparaison sur la période de prévision des modèles :

Après avoir calculé les paramètres des modèles sur la période de calibration de 12 mois, nous avons utilisé et comparé les modèles Petri et NBD avec régression en terme de prévision.

Nous avons par conséquent simulé, à partir du réseau de Petri(au niveau individuel), une période de 4 mois, soit les périodes 13-16 mois après la date de création de la première liste. Le choix de la longueur de cette période est lié à la période d'historique maximum que nous avions sur un maximum de listes (rappelons que cette période est de 16 mois. Les 12 premiers mois ont servi pour la calibration, les 4 derniers pour la prévision).

Les résultats sur la variable nombre d'items achetés sont les suivants :

Tableau 4.14: résultats NBD et Pétri pour la prévision

	Nombre d'items achetés			Différence Observé-Modèle	
	Observés	NBD	Pétri	NBD	Pétri
130961	0	0	0,74	0	-0,74
akwell1	0	0,18	5,28	-0,18	-5,28
AlexO	1	0,66	3,65	0,34	-2,65
ama1	1	0,29	0	0,71	1
amibio	0	0,06	0	-0,06	0
AnnA	4	2,71	3,56	1,29	0,44
Aude	1	0,29	1,92	0,71	-0,92
badir	0	0,14	0	-0,14	0
bgougay	0	0,39	0,76	-0,39	-0,76
blondi	1	0,25	0	0,75	1
bpdl	0	0,36	1,8	-0,36	-1,8
bruyere	0	0,33	0	-0,33	0
bschunke	1	0,13	0,26	0,87	0,74
carolala	0	0,24	0	-0,24	0
cathou1	0	0	0	0	0
cclerget	0	0	0	0	0
Céline21	0	0,39	3,25	-0,39	-3,25
chap's	0	0,19	0,87	-0,19	-0,87
Cypher	6	0,25	0,28	5,75	5,72
DALMATE	0	0,06	0	-0,06	0
Delatte	1	0,26	0,71	0,74	0,29
didine77	0	0	0	0	0
Dinou	0	0,1	1,19	-0,1	-1,19
djoule	11	2,91	8,2	8,09	2,8
dolores3	1	0,12	0	0,88	1
doppel	0	0,12	0,04	-0,12	-0,04
eb95	1	0,21	0	0,79	1
eb96	0	0,12	0,01	-0,12	-0,01
Emilie	0	0,15	0	-0,15	0
estia	3	0,12	0	2,88	3
eusebio1	3	0,33	0	2,67	3

Tableau 4.14: résultats NBD et Pétri pour la prévision

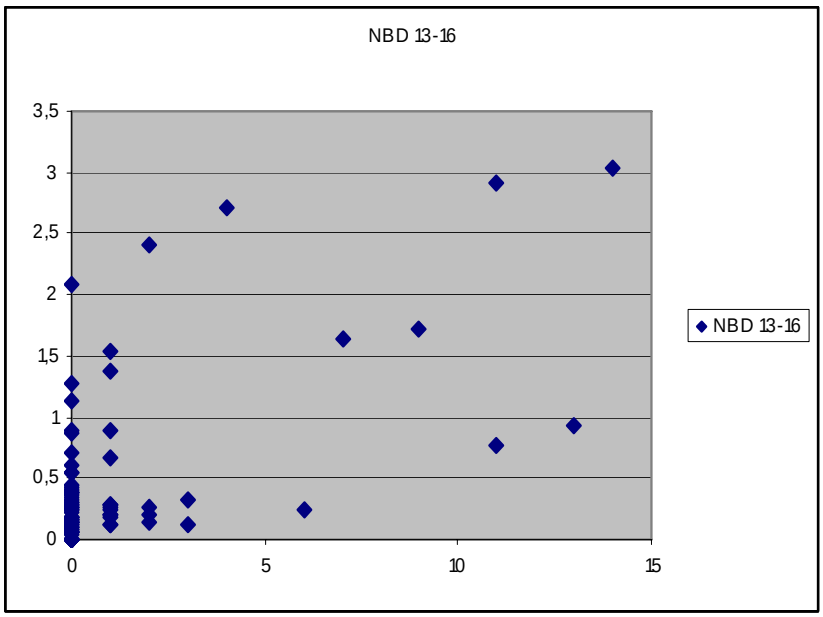
	Nombre d'items achetés			Différence Observé-Modèle	
	Observés	**NBD**	**Pétri**	**NBD**	**Pétri**
fabou	0	0,06	0	-0,06	0
fallouis	0	0,06	0,27	-0,06	-0,27
fevrard	0	0	0,05	0	-0,05
FOURNIAL	1	0,19	2,35	0,81	-1,35
Fruor	0	0	2,38	0	-2,38
frzpower	2	0,15	0	1,85	2
GaëlleM	2	0,26	0,65	1,74	1,35
gallouin	13	0,94	0,99	12,06	12,01
Ginie31	2	2,41	0	-0,41	2
Greensky	0	0,3	1,39	-0,3	-1,39
gzing	0	0,19	0	-0,19	0
HERPIN	0	0,24	0	-0,24	0
hutch	2	0,2	0	1,8	2
jeanbask	1	0,21	0	0,79	1
jeannot3	9	1,71	4,34	7,29	4,66
jeda	0	0,14	0	-0,14	0
JIMASA	0	0	0	0	0
joewj	0	0,38	0	-0,38	0
juboubou	0	0,11	0	-0,11	0
juju75	0	0	0	0	0
kabotine	11	0,77	0	10,23	11
Kifran	0	1,13	0	-1,13	0
lbarbier	0	1,27	0	-1,27	0
levieux	0	0,55	0	-0,55	0
Loquet	0	0,54	0	-0,54	0
luberdov	0	0	0	0	0
m@hy	0	0,42	0	-0,42	0
manwe	0	0,27	0	-0,27	0
mary_ce	0	0,15	0	-0,15	0
Mayfair	0	0,34	0	-0,34	0
mekiela	0	0	2,14	0	-2,14
Michel	0	0,11	0	-0,11	0
minious	0	0,41	0	-0,41	0
mperrot	0	0,6	1,41	-0,6	-1,41
Munier	0	0,7	0	-0,7	0
Naemesis	0	0,86	0	-0,86	0
nhai07	0	0,29	0	-0,29	0
ninie12	0	0,09	0	-0,09	0
philia	0	0,89	0,39	-0,89	-0,39
picardjf	0	0,31	0	-0,31	0
popcorn	0	0,07	0	-0,07	0
prioton	0	0,45	0	-0,45	0
prozaiik	0	2,09	0	-2,09	0
quentin	0	0,19	0	-0,19	0
RAPH	0	0,17	0	-0,17	0
rlaval	1	0,89	0,67	0,11	0,33

Tableau 4.14: résultats NBD et Pétri pour la prévision

	Nombre d'items achetés			Différence Observé-Modèle	
	Observés	NBD	Pétri	NBD	Pétri
romain37	0	0,27	0	-0,27	0
sandor	0	0	0	0	0
semery	14	3,04	1,64	10,96	12,36
sophynet	0	0,15	0	-0,15	0
Souris	0	0,09	0	-0,09	0
starchen	0	0	0	0	0
Stèph	0	0	0	0	0
stumpf	0	0	0	0	0
Sylvie4	0	0,16	0	-0,16	0
Thibaut	1	1,38	2,48	-0,38	-1,48
toto2	0	0,28	0	-0,28	0
tranb	0	0,16	0	-0,16	0
vianney2	0	0,22	1,41	-0,22	-1,41
Virginie	0	0,27	0	-0,27	0
xzouf	1	1,54	0	-0,54	1
Yseulys	7	1,63	1,24	5,37	5,76
Yves	0	0,05	0	-0,05	0

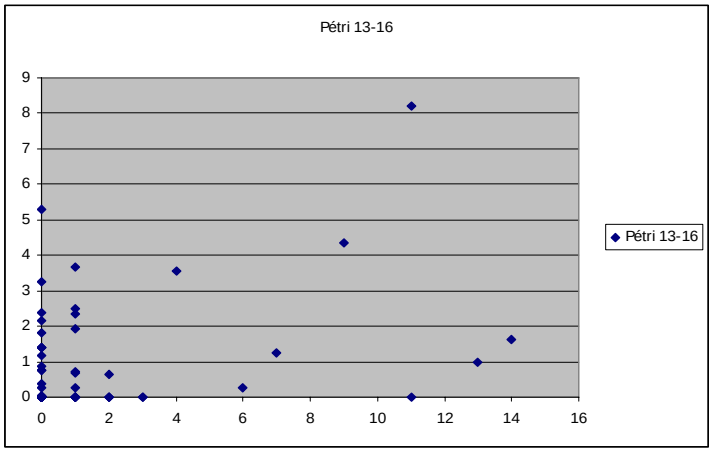
Une synthèse de ce tableau est présentée ci-dessous sous forme de graphique comparant les valeurs estimées par le modèle (soit NBD soit Pétri) en ordonnées et les valeurs observées en abscisse, pour chacun des agents.

Nbre d'items achetés estimés par le modèle NBD
versus données observées



Nbre d'items achetés estimés par le modèle Pétri

versus données observées



Le tableau 4.14 montre le nombre d'items observé, simulé par le modèle NBD avec régression, simulé par Petri, puis les différences entre les données observées et chacun des deux modèles. Lorsqu'il n'y a aucun achat, les deux modèles sont proches et donnent également un nombre d'items achetés proche de zéro. Dans quelques cas isolés, les modèles divergent fortement : pour le login gallouin ou semery par exemple, aucun des deux modèles ne donne de bons résultats.

Dans 84 cas sur les 94 présentés dans le tableau 4.14, le modèle de Petri est plus performant sur le critère MAD (différence entre données observées et modèle).

Une synthèse du tableau 4.14 est fournie ci-dessous lorsqu'on synthétise ces écarts sous les deux indicateurs MAD et MSE présentés précédemment.

MSE

13-16 months	NBD regression	7,46
	PETRI	6,65

MAD

13-16 months	NBD regression	0,75
	PETRI	0,49

TABLEAU 4.15

Le tableau 4.15 permet de conclure à une meilleure performance du modèle de Petri par rapport au modèle NBD également sur la partie prévision. Quelques nuances peuvent

cependant être soulignées. Tout d'abord, on constate que dans le cadre de la prévision, les réseaux de Petri sont plus proches du modèle NBD que dans le cadre de la calibration des modèles pour laquelle nous avions également comparé les deux modèles. Ceci s'explique essentiellement par le fait que la période de prévision, en raison des données disponibles que nous avions, est particulièrement courte (4 mois), et que sur cette période de 4 mois, de nombreuses listes se trouvent sans achat (dans ce cas les deux modèles sont équivalents).
De plus, la comparaison menée ici entre les réseaux de Petri et le modèle NBD avec régression a été effectuée avec un modèle de Petri par agent A. Par conséquent, le nombre de paramètres à estimer est important (6 paramètres par login, à savoir les 2 paramètres de la loi de connexion des agents A, les 2 paramètres de la loi de connexion des agents B, les 2 paramètres de la loi NBD d'achat) alors que le modèle NBD avec régression a 4 paramètres (2 paramètres de la loi NBD et 2 coefficient de régression pour les variables connexions et age).

Dans cette partie 4.2, nous avons souligné les bonnes performances du modèle de Petri par rapport au modèle NBD avec régression utilisé plus classiquement en marketing pour mesurer le nombre de produits achetés par un consommateur. Nous avons également mis en avant la flexibilité des réseaux de Petri qui permettent de modéliser des actions séquentielles (connexion puis achat), des actions parallèles (connexions agent A et connexions agent B) en déterminant les lois de chacun des évènements survenant dans le système.
Nous allons dans la partie suivante étendre le modèle des wish lists et utiliser les réseaux de Petri pour mesurer l'impact de l'e-mailing sur le nombre d'items achetés.

4.3 EXTENSION A LA MESURE DE L'IMPACT DE L'E-MAILING ET IMPLICATIONS MANAGERIALES:

Dans le cadre de la gestion des wish lists et afin d'améliorer le trafic et les achats sur les listes de cadeaux, la société Millemercis envoie régulièrement des e-mails aux créateurs de la liste (que nous avons nommés agents A). Ces e-mails sont des e-mails publicitaires émis par Millemercis en collaboration avec l'un de ses partenaires. Un extrait de la liste de ces e-mails est fourni ci-dessous à titre d'exemple (ce sont les e-mails envoyés durant les trois premiers mois de l'année 2002). On constate que des secteurs très divers constituent les supports de ces e-mails, de la banque (Crédit Agricole, Crédit Lyonnais) à Auchan, Cegetel, Castorama,

Bayard presse ou Toyota. On peut noter également la fréquence rapprochée de ces e-mails (même si certains ont été ciblés).

DESCRIPTIF MAIL	DATE MAILING
Aucland MAIL1	08/01/2002
Aucland MAIL2	08/01/2002
Toyota1	10/01/2002
Credit lyonnais etudiant	14/01/2002
Credit lyonnais jeunes actifs	16/01/2002
Credit lyonnais investisseurs	28/01/2002
Toyota2	10/01/2001
Castorama (routage externe)	06/12/2001
Bayard Presse (routage externe)	30/01/2002
Cmescourses (routage externe)	31/01/2002
Nocibé	04/02/2002
Education.com	04/02/2002
News PcCity 2	04/02/2002
Sélectionclic	04/02/2002
PC City	07/02/2002
Pc City we camescope16-17fev	14/02/2002
Pc City we camescope 16-17fev bis	15/02/2002
Meetic1	18/02/2002
Meetic2	19/02/2002
Meetic3	20/02/2002
Meetic3	20/02/2002
Pc City sony packard bell	20/02/2002
Haagen Dazs Mystère test 500	25/02/2002
Haagen Dazs test 500	25/02/2002
Haagen Dazs	26/02/2002
Meetic4	04/03/2002
Meetic5	04/03/2002
NetNoe/alysse sur le dépt 92	05/03/2002
Crédit Agricole	06/03/2002
Telemarket	12/03/2002
Meetic H	11/03/2002
Meetic F	11/03/2002
Castorama	14/03/2002
education.com	19/03/2002
Meetic H2	19/03/2002
Meetic F2	19/03/2002
Wanadoo	21/03/2002
Cegetel	21/03/2002
Bayard Web	21/03/2002
Auchan Direct	27/03/2002
AOL	29/03/2002
Castorama Relance	27/03/2002

4.3.1 Intérêt de modéliser l'e-mailing :

Durant les dix dernières années, les responsables marketing et les chercheurs se sont intéressés aux stratégies sur Internet. Leurs contributions concernent majoritairement l'amélioration du design des sites, la mesure de sa performance, la mesure d'audience, la satisfaction d'un consommateur de sites web et plus globalement le comportement de l'internaute. Cependant, l'action la plus fréquente d'un internaute consiste à ouvrir sa boîte e-mail, à consulter ses messages, à recevoir et envoyer des e-mails. En 2004, le nombre de boîtes e-mails dans le monde a dépassé le milliard (source : messagingonline) et l'e-mail est devenu un moyen de communication fondamental, à la fois sur le plan professionnel et personnel. Il constitue pour les entreprises un outil marketing puissant et peu onéreux (Chittiden et Rettie, 2003), en comparaison avec d'autres médias plus classiquement utilisés en marketing, comme la télévision ou la radio par exemple. Malgré cette tendance lourde, peu de recherches en marketing portent sur l'e-mail. Quelques articles récents se concentrent sur le taux de réponse aux e-mails et les facteurs influençant ce taux de réponse (Ken et Brandal, 2003 ; Chittiden et Rettie, 2003 ; Sheehan &McMillan, 1999). Ces articles en revanche n'analysent pas l'impact de l'e-mail sur le nombre de connexions ou d'achats sur le site Web. Anasari et Mela (2003) ont modélisé la « customisation » des e-mails à partir des données de navigations des internautes. Ils montrent notamment que l'optimisation du contenu des e-mails d'une campagne opt-in (l'internaute a accepté au préalable de recevoir les e-mails) peut améliorer le taux de click des e-mails de 62%. En revanche, ces auteurs n'ont pas effectué de lien avec le nombre de connexions et le nombre de produits achetés.

L'objectif de cette partie est donc d'essayer de mesurer l'impact des e-mails publicitaires envoyés par Millemercis aux agents A. Cet impact sera mesuré par le nombre de connexions des agents A sur le site et par le nombre d'items achetés. Plus précisément, nous souhaitons analyser l'impact de la fréquence d'envoi des mails. En effet, de nombreuses entreprises aujourd'hui utilisent l'e-mail dans leur campagne de communication mais peu connaissent la fréquence d'envoi optimum ou tout au moins la meilleure fréquence à adopter parmi plusieurs options possibles. Nous allons donc nous intéresser plus particulièrement à l'impact de la durée inter e-mail sur les performances des wish lists en terme de connexions et d'achats.

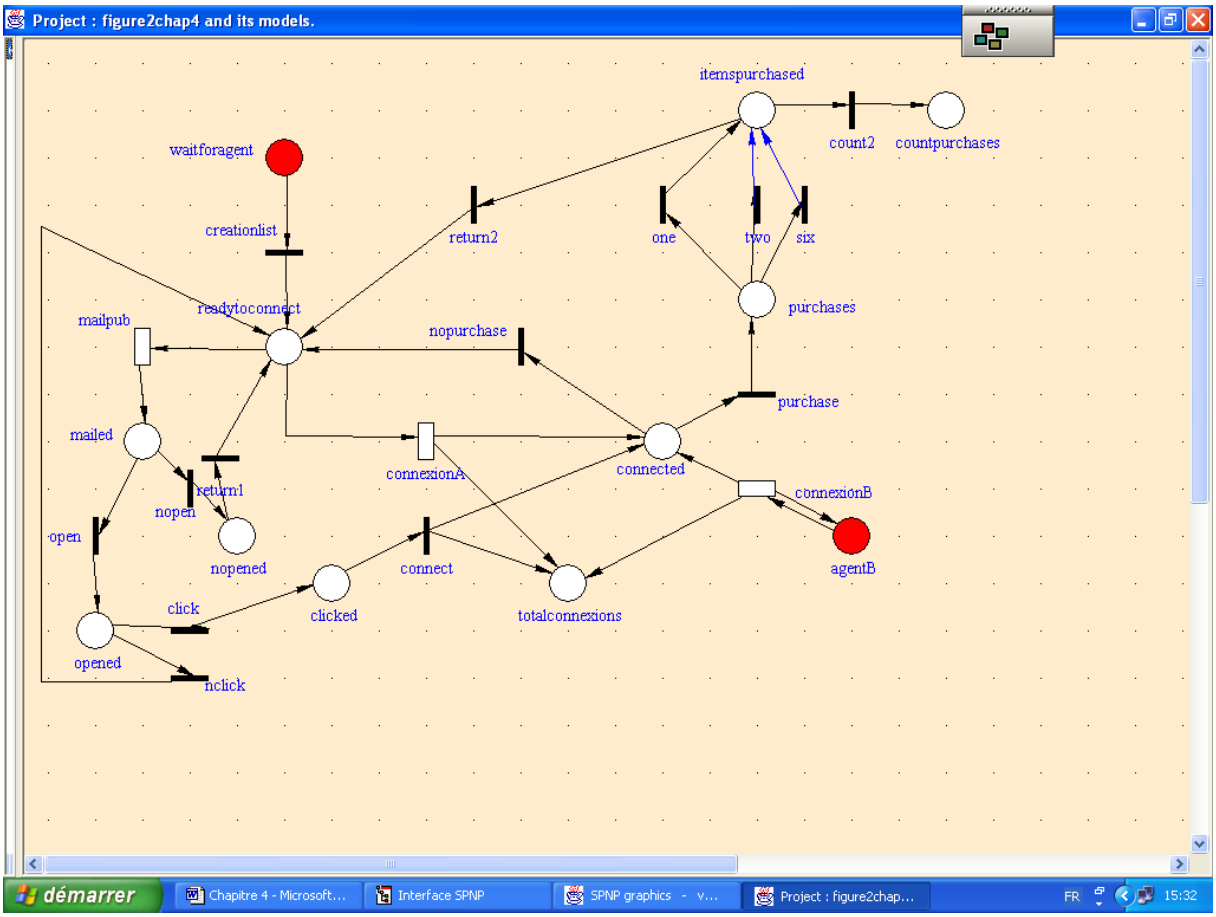
4.3.2 Modélisation de l'e-mailing par les réseaux de Petri:

Afin d'analyser l'impact des e-mails, nous avons intégré dans le réseau de Petri élaboré pour chaque agent A un certain nombre de places et de transitions décrivant le phénomène d'e-mailing.

Ce système est présenté dans la figure 4.9 ci-dessous sur un agent A choisi à titre d'exemple : l'agent akwell1.

Figure 4.9: login akwell1

Wish list incluant l'e-mailing modélisée par un réseau de Petri



La figure 4.9 montre que nous avons intégré dans la partie gauche du graphique le phénomène d'e-mailing. Ceci se traduit dans le réseau de Petri par la création de la transition *mailpub*, représentant l'événement selon lequel l'entreprise envoie à l'agent A un e-mail publicitaire tel que décrit précédemment. Une fois l'e-mail envoyé, le système se trouve dans la place *mailed* (un mail a été envoyé à l'agent A). Le mail peut ensuite être soit ouvert (transition *open*) ou non ouvert par l'agent A (transition *nopen*). Le cas où l'agent n'ouvre pas l'e-mail est

représenté dans le système par la place *nopened*, suivie ensuite de la transition *return1* indiquant que le système revient à son état initial dans lequel l'agent A est prêt à se connecter sur le site (soit sous l'impulsion d'un prochain mail reçu, soit par sa propre démarche).
Si l'agent A en revanche ouvre l'e-mail, le système évolue dans l'état *opened*. Cet état est suivi de soit un click sur l'e-mail (transition *click*), soit aucun click sur l'e-mail (transition *nclick*). Dans le cas où l'agent A ne clique pas sur l'e-mail, on revient également à l'état initial dans lequel l'agent A peut se connecter à nouveau. Si l'agent A clique sur l'e-mail, on passe à l'état *clicked*, suivi de la transition *connect* puisqu'automatiquement le click sur le lien contenu dans l'e-mail envoit l'internaute vers le site www.Millemercis.com. L'agent A est alors connecté (place *connected*) et il peut soit acheter soit ne pas acheter d'items selon la description du réseau de Petri sans e-mail que nous avons effectuée au chapitre 3. Notons que les transitions *click* et *nclik* sont immédiates ce qui signifie qu'elle s'effectuent dès qu'il y a un jeton dans la place *opened* (donc dès qu'un e-mail a été ouvert). Nous ne tenons pas compte ici du temps entre l'ouverture de l'e-mail et le click sur celui-ci.

4.3.3 Modèle HLM pour déterminer le taux d'ouverture d'un e-mail :

Nous nous sommes donc intéressée dans un premier temps à la modélisation du taux d'ouverture d'un e-mail (probabilité de franchissement de la transition *open* et par voie de conséquence de la transition nopen) en fonction d'un certain nombre de variables. En effet, un individu reçoit plusieurs e-mails et décide de les ouvrir ou non. Cet ouverture peut être influencée par deux types de caractéristiques : le type d'e-mail envoyé et les caractéristiques propres à l'individu qui le reçoit. Etant donné ces deux niveaux d'influence, nous avons opté pour un modèle hiérarchique linéaire pour modéliser le taux d'ouverture des e-mails.
Le modèle HLM (Hierarchical Linear Model) à deux niveaux, avec un lien logit pour des résultats binaires, développé par Raudenbush et Bryk (2002), paraît l'un des meilleurs modèles pour expliquer le phénomène étudié. Les modèles HLM sont utilisés lorsque les données comportent plusieurs niveaux (par exemple un niveau individuel (niveau 1) et un niveau groupe (niveau 2)). Le principe de cette méthode est le suivant : le modèle HLM effectue tout d'abord une régression linéaire sur le niveau 1 et utilise les variables du niveau 2 pour expliquer les variations des paramètres au niveau individuel.

Nous considérons ici un modèle HLM à 2 niveaux : le niveau 1 correspond aux caractéristiques de l'e-mail, le niveau 2 au profil de l'individu, ces deux niveaux influençant la probabilité d'ouvrir l'e-mail reçu.
Ce modèle décrit la probabilité d'ouvrir un e-mail i par l'individu j comme une fonction du profil de l'individu et du type d'e-mail :

prob(ouvrir e-mail$_{ij}$) = f(profil individu$_j$, profil e-mail$_{ij}$)

Niveau 1 du modèle HLM :

Soit Y_{ij} le nombre des e-mails ouverts pour m_{ij} e-mails envoyés, et soit φ_{ij} la probabilité d'ouvrir à chaque envoi. Nous écrivons alors :

$$Y_{ij} \left| \varphi_{ij} \right. \sim B(m_{ij}, \varphi_{ij})$$

ce qui signifie que Y_{ij} suit une distribution binomiale de paramètres m_{ij} (nombre d'envois d'e-mails) et φ_{ij} (probabilité d'ouverture à chaque envoi). Conformément à la distribution binomiale, l'espérance et la variance de Y_{ij} sont :

$$E(Y_{ij} \left| \varphi_{ij} \right.) = m_{ij}\varphi_{ij}, \qquad Var(Y_{ij} \left| \varphi_{ij} \right.) = m_{ij}\varphi_{ij}(1-\varphi_{ij}).$$

Fonction de lien du niveau- 1 :

$$\eta_{ij} = \log\left(\frac{\varphi_{ij}}{1-\varphi_{ij}}\right),$$

où η_{ij} = log-odds d'ouverture de l'e-mail.

Plusieurs fonctions de lien entre η_{ij} et φ_{ij}, comme probit, log-normal ou logit sont possibles lorsque le modèle est binomial (Hedeker and Gibbons 1994). Nous avons ici choisi le plus connu et le plus simple, le lien logit.

Lorsque la probabilité φ_{ij} d'ouverture de l'e-mail est inférieure à 0.50, $\frac{\varphi_{ij}}{1-\varphi_{ij}} < 1$ et le logarithme est négatif. Lorsque la probabilité d'ouverture φ_{ij} de l'e-mail est supérieure à 0.50, $\frac{\varphi_{ij}}{1-\varphi_{ij}})$ odds > 1 et le logarithme est positif.

Il faut noter que φ_{ij} se trouve dans l'intervalle [0,1] mais η_{ij} peut prendre toute valeur d'un nombre réel. .

Modèle structurel niveau –1 :

Les valeurs prédites η_{ij} sont reliées aux variables X_{ij} (caractéristiques de l'e-mail i pour l'individu j) par le modèle structurel linéaire suivant:

$$\eta_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + ... + \beta_{pj} X_{p1ij} .$$

et $$\varphi_{ij} = \frac{1}{1 + \exp(-\eta_{ij})}$$

Quelle que soit la valeur de η_{ij}, l'équation ci-dessus produira une valeur de φ_{ij} entre zéro et un comme probabilité d'ouverture d'un e-mail.

Modèle niveau-2

Le modèle niveau-2 permet d'intégrer le profil de l'individu dans le modèle. Il a la forme suivante:

$$\beta_{qj} = \gamma_{q0} + \sum_{s=1}^{S_q} \gamma_{qs} W_{sj} + u_{qj} ,$$

avec :

- les termes d'erreur u_{qj} ($q=0,...,Q$) constituent un vecteur uj dont la distribution suit une loi normale multivariée
- les variables W_{sj} correspondent aux caractéristiques de l'individu j

Sur les 135 agents A, 12 ont été enlevés pour l'estimation du modèle HLM en raison de données de profil manquantes. En moyenne, les agents A ont reçu un e-mail chaque 23 jours. On notera INT la variable intervalle de temps entre deux e-mails (dans le modèle sera intégré

la variable logarithme de l'intervalle, notée 'LN INT'). Le taux d'ouverture moyen d'un e-mail est de 54% (variable 'OPEN'). En moyenne, les individus se sont connectés sur leur liste de cadeau 4,75 fois entre deux e-mails (variable 'CONNECT') et leur taux d'ouverture du précédent e-mail reçu est de 52% (variable 'OPENP'). La moyenne d'âge est de 29 ans (variable 'AGE') et les individus ont créé en moyenne 1,6 liste de cadeaux (la variable nombre de liste est notée 'LIST'). 47% des individus sont des hommes (la variable sexe est notée 'GENDER') et 56% ont fourni leur numéro de portable (variable 'CELLBIN') à l'entreprise Millemercis.

Les résultats sur l'ensemble des 123 agents A traités ensemble dans le modèle HLM sont les suivants :

Tableau 4.16: Résultats du modèle HLM sur l'ensemble des agents A

	Coefficient	S.E.	T-ratio	Approx d.f.	P-value
Level 1					
OPENP (ouverture mail précédent)	1,271	0,116	10,98	122	0
CONNECT (nbre de connexions entre 2 e-mails)	-0,002	0,004	-0,451	122	0,652
LN INT (LN intervalle entre 2 e-mails)	-0,326	0,209	-1,562	122	0,118
LN INT2 (LN (intervalle entre 2 e-mails)^2)	0,045	0,036	1,242	122	0,215
Level 2					
CONSTANTE	0,056	0,443	0,127	118	0,899
AGE	-0,014	0,008	-1,818	118	0,069
LIST (Nbre de listes)	0,144	0,052	2,76	118	0,006
CELLBIN (no de portable fourni)	0,288	0,163	1,768	118	0,077
GENDER (sexe)	0,122	0,159	0,77	118	0,441

Le tableau 4.16 montre que sur l'ensemble des 123 agents, l'effet de l'intervalle entre deux e-mails se révèle non significatif (les p-values des variables LN INT et $(LN\ INT)^2$ sont supérieures à 0,05. Le niveau 1 fournit les variables liées aux caractéristiques de l'e-mail et explicatives de la probabilité d'ouverture d'un e-mail. Le niveau 2 fournit les variables liées aux caractéristiques de l'individu et explicatives de la probabilité d'ouverture d'un e-mail.

Les seules variables significatives au niveau–2 sont l'âge, le nombre de listes et le fait d'avoir fourni son numéro de portable, au seuil significatif α=10%).

Ces résultats peuvent s'expliquer par le fait que les données agrégées sur l'ensemble des 123 agents cachent une hétérogénéité du comportement d'ouverture d'un e-mail qui n'est pas bien capturé par le modèle ci-dessus.

Afin de résoudre ce problème, nous introduisons une hétérogénéité en divisant les agents en deux groupes selon leur comportement d'ouverture des e-mails.

Pour cela, nous avons mené une régression (par la méthode des moindres carrés ordinaires) de la variable taux d'ouverture en fonction de deux variables liées à l'intervalle entre deux e-mails (LN INT, logarithme de l'intervalle ; LN INT^2, logarithme du carré de l'intervalle). Seuls les agents dont le taux d'ouverture suit une courbe en U inversée (coefficient positif de LN INT et négatif de LN INT^2) sont sélectionnés dans le groupe 1. Le reste des agents constitue le groupe 2. Les résultats de cette régression ainsi que la classification par groupe (1 et 2) sont fournis dans le tableau ci-dessous pour chaque login (id dans le tableau).

ID	Constante	OPEN SLOPE	CONNECT SLOPE	LNTIMESP SLOPE	LNTIIME2 SLOPE	Group
1	0,13612	0,70739	-0,00846	2,49	-0,44	1
7	0,74849	-0,24789	-0,19954	1,48	-0,23	1
9	0,07424	0,85152	-0,04257	3,87	-0,57	1
12	0,38957	0,10237	0,01543	0,57	-0,09	1
26	0,8887	-0,09856	-0,19465	0,05	-0,02	1
27	0,97583	-0,05715	0,00544	0,63	-0,11	1
29	0,26338	0,33262	0,08285	1,21	-0,22	1
31	0,07983	-0,01891	0,01694	0,21	-0,06	1
42	0,51813	-0,04532	0,02214	0,08	-0,01	1
44	0,39132	0,23909	-0,02064	0,29	-0,06	1
46	0,33721	0,22674	0,04395	0,51	-0,08	1
49	0,13938	-0,10283	0,04049	0,77	-0,11	1
52	0,32109	-0,28435	8,80753	18,20	-3,26	1
54	1,02245	-0,14274	0,03197	0,81	-0,13	1
57	0,76164	-0,17707	0,02375	0,86	-0,10	1
58	0,63082	-0,18781	-0,01272	4,43	-0,68	1
61	0,49413	0,13528	0,06614	0,61	-0,11	1
62	0,51433	0,17706	0,02631	5,38	-0,79	1
66	0,38381	0,53263	-0,00165	0,60	-0,11	1
68	1,10328	-0,26089	0,05661	0,45	-0,04	1
69	0,31769	-0,14367	-0,05202	1,01	-0,13	1
71	0,18721	0,10503	-0,13184	2,07	-0,27	1
77	0,50157	0,49819	-0,0268	0,61	-0,10	1
79	0,46733	0,08168	0,07262	2,60	-0,44	1
90	0,11348	-0,13481	0,00096	0,12	-0,02	1
91	0,22112	0,40198	-0,35293	2,91	-0,45	1
104	1,12388	-0,32822	0,06848	12,20	-1,78	1
106	0,45523	0,01367	0,13708	0,48	-0,09	1
107	0,06491	-0,10354	-0,02047	0,31	-0,05	1
110	0,05995	-0,06955	-0,00833	0,50	-0,09	1
112	0,24579	0,38552	0,0019	0,27	-0,02	1
117	0,49561	0,32456	-0,00039	0,41	-0,05	1
126	0,45739	0,19733	-0,01339	0,06	-0,05	1
3	0,20042	0,34178	0,00415	-1,19	0,18	2
4	0,2468	0,75375	-0,0063	-0,01	0,00	2
5	0,2209	0,5582	0,13496	-0,47	0,02	2
6	0,48663	0,20756	-0,00355	-0,37	0,06	2
8	0,49741	0,41701	-0,00251	-0,26	0,06	2
11	0,38447	0,11571	0,00522	-0,47	0,05	2
13	0,37688	0,54272	0,02746	-0,10	0,01	2
15	0,22516	0,86646	0,20076	-3,72	0,59	2
16	0,42112	0,2236	0,05103	-0,07	-0,05	2
18	0,43742	0,05374	-0,11806	0,18	0,00	2
19	0,51264	0,16696	-0,05744	-0,20	0,06	2
20	0,54992	0,25012	0,01474	-1,23	0,21	2
21	0,5436	-0,02054	-0,13202	-0,22	-0,00	2
22	0,92744	-0,39117	0,24851	-0,88	0,06	2
25	0,18372	-0,04108	0,11622	-2,24	0,32	2
28	0,61075	0,06172	0,01106	-1,00	0,16	2
30	0,43241	0,37206	0,07709	-0,13	-0,04	2
32	0,03653	0,97038	-0,00404	-0,00	0,01	2
33	0,22719	0,71602	-0,07986	-0,03	0,00	2
34	0,66129	0,09467	-0,01044	-0,80	0,21	2
35	0,11301	0,32536	-0,08425	-0,38	0,06	2
36	0,21161	0,08615	0,0955	-0,83	0,12	2
39	0,87857	-0,43	0,17687	-5,09	0,69	2
40	0,60051	0,24284	-0,03214	-0,18	0,03	2
41	0,33077	0,47155	-0,00316	-0,00	0,01	2
48	0,08817	-0,05804	0,06893	-2,01	0,34	2
50	0,52933	-0,05028	1,01311	-0,07	-0,08	2
51	0,17836	0,61085	0,09798	-0,17	0,04	2
56	0,51584	0,18712	0,0008	-0,31	0,02	2
59	0,49039	0,24388	-0,00152	-0,17	0,03	2
64	0,49482	0,12032	-0,22356	-2,06	0,28	2
67	0,49864	0,45616	0,00095	-0,01	-0,00	2
70	0,39354	-0,04944	0,02109	-0,76	0,16	2
72	0,51216	0,21019	0,01827	0,22	0,00	2
74	0,36562	0,53692	0,0008	-0,46	0,08	2
76	0,20351	0,79539	-0,0012	-0,06	0,00	2
83	0,37498	-0,31242	-0,01633	-2,75	0,36	2
85	0,71074	-0,56612	-0,38803	-5,28	0,75	2
87	0,29067	0,14606	-0,01026	-0,22	0,04	2
89	0,60191	0,16428	0,11594	-0,16	0,02	2
94	0,2612	0,43011	0,0134	-0,09	0,02	2

D'autre part, la distribution de l'intervalle entre deux e-mails, sur l'ensemble des données observées (2206 e-mails envoyés), est la suivante :

Tableau 4.17 : Statistiques sur l'intervalle entre 2 e-mails

Statistics

TIMESPAN

N	Valid	2206
Percentiles	5	3,0000
	10	7,0000
	15	9,0000
	20	11,0000
	25	14,0000
	30	16,0000
	35	19,0000
	40	20,0000
	45	22,0000
	50	25,0000
	55	26,0000
	60	28,0000
	65	31,0000
	70	33,0000
	75	38,0000
	80	43,0000
	85	50,0000
	90	58,0000
	95	78,0000

Le tableau 4.17 montre que 80% des intervalles entre e-mails se situent entre 7 et 58 jours. Les intervalles inférieurs à 7 jours constituent 10% des données observées, ceux supérieurs à 58 jours 10% également des données observées.

Nous allons donc étudier pour chaque groupe 1 et 2 le comportement d'ouverture des e-mails en fonction de l'intervalle entre e-mails, puis nous analyserons les résultats du modèle HLM appliqué à chacun des 2 groupes.

Les résultats du groupe 1 sont les suivants :

La probabilité d'ouvrir un e-mail dans le groupe 1 (33 agents A) en fonction de l'intervalle entre deux e-mails suit donc une courbe en U inversée, comme le montre la figure 4.10 ci-dessous.

Figure 4.10: résultats groupe 1

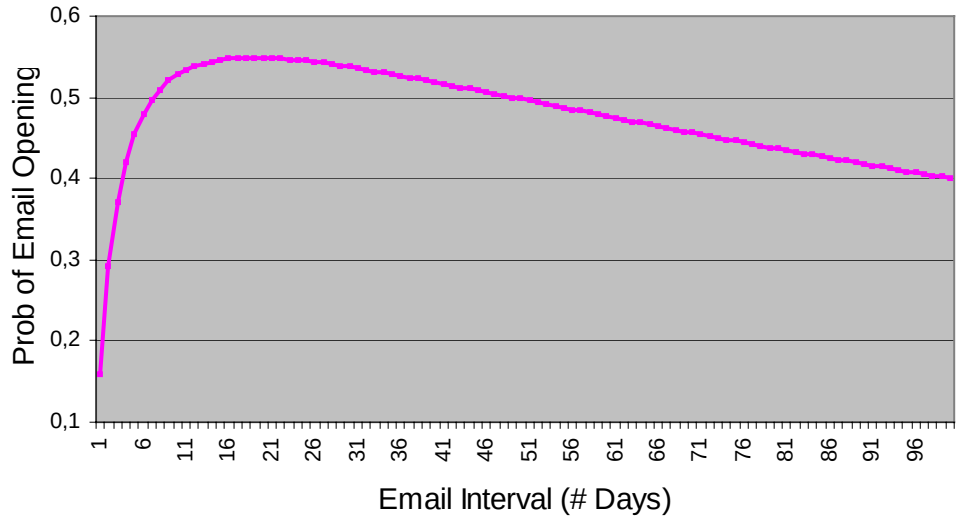


Tableau 4.18 : résultats du modèle HLM groupe 1

	Coefficient	S.E.	T-ratio	Approx d.f.	P-value
Level 1					
OPENP	0,818	0,207	3,952	32	**0,001**
CONNECT	0,004	0,007	0,535	32	0,596
LN INT	1,264	0,455	2,775	32	**0,010**
LN INT2	-0,215	0,079	-2,714	32	**0,011**
Level 2					
INTCEPT	-1,656	0,855	-1,937	28	**0,062**
AGE	-0,014	0,013	-1,073	28	0,293
LIST	-0,068	0,158	-0,429	28	0,671
CELLBIN	0,441	0,358	1,23	28	0,229
GENDER	-0,273	0,341	-0,803	28	0,429

Dans le groupe 1, l'effet de l'intervalle entre deux e-mails (en forme de U inversé) devient significatif (au taux d'erreur α=5%) avec un effet positif de l'ouverture du mail précédent (coefficient 0.818), alors qu'il ne l'était pas en considérant l'ensemble des individus. Donc plus l'agent a ouvert l'e-mail précédent, plus sa probabilité d'ouverture de l'e-mail est importante. En ce qui concerne l'intervalle, les coefficients sont de 1.264 pour le log de l'intervalle et de –0.215 pourle log de l'intervalle au carré. On a donc une courbe en U

inversée avec une probabilité d'ouvrir un e-mail qui augmente lorsque l'intervalle entre 2 e-mails augmente puis qui décroît après un pic (ce pic correspond à un intervalle de 18.9 jours entre l'e-mail reçu et le précédent), comme on peut le constater dans ma figure 4.10.
On peut par conséquent sur ce groupe proposer comme recommandation managériale l'envoi d'un e-mail tous les 18,9 jours dans l'optique de maximiser le taux d'ouverture (figure 4.10).
En revanche, il n'y a pas pour ce groupe 1 d'impact au niveau individuel des caractéristiques liées à l'agent (les variables 'AGE', nombre de listes ('LIST'), fourniture du numéro de portable ('CELLBIN'), et sexe ('GENDER') ne sont pas statistiquement significatives (dans le tableau 4.18, les p-value >0.10 pour ces variables).

Les résultats du groupe 2 sont les suivants

Figure 4.11: résultats groupe 2

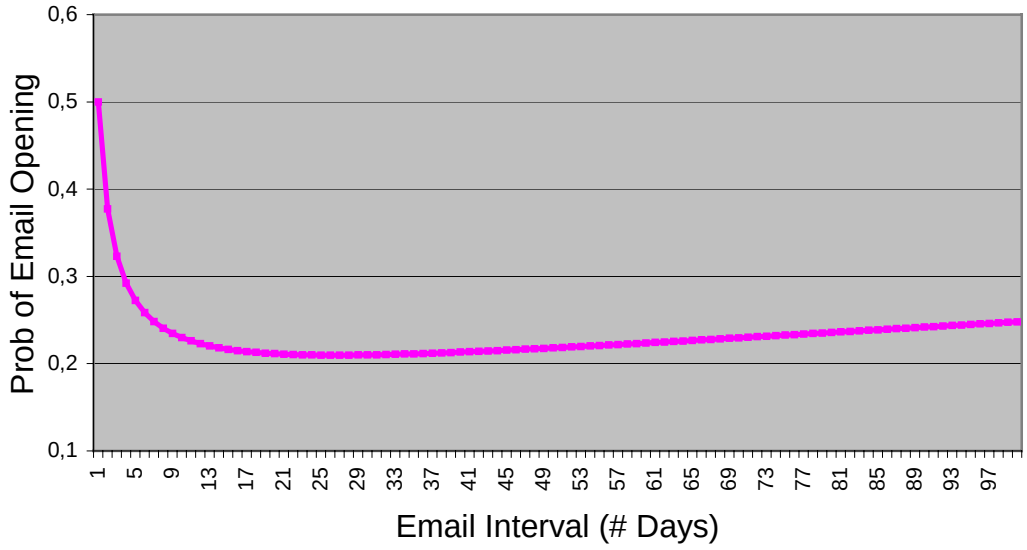


Tableau 4.19 : résultats du modèle HLM groupe 2

	Coefficient	S.E.	T-ratio	Approx d.f.	P-value
Level 1					
OPENP	1,456	0,156	9,337	89	**0,000**
CONNECT	-0,004	0,005	-0,9	89	0,369
LN INT	-0,809	0,242	-3,347	89	**0,001**
LN INT2	0,123	0,042	2,952	89	**0,004**
Level 2					
INTCEPT	0,518	0,540	0,96	85	0,337
AGE	-0,012	0,010	-1,239	85	0,216
LIST	0,166	0,057	2,92	85	**0,004**
CELLBIN	0,314	0,184	1,71	85	**0,087**
GENDER	0,240	0,180	1,332	85	0,183

Pour le groupe 2 (90 agents), l'effet de l'intervalle entre deux e-mails suit une courbe en U comme le montre la figure 4.11. Ceci signifie que la probabilité d'ouverture décroît en fonction de l'intervalle jusqu'à un minimum (pour un intervalle de 28 jours) puis augmente à nouveau lorsque l'intervalle est supérieur à 28 jours. On constate donc ici un comportement différent de celui des agents du groupe 1.

Dans les données utilisées, les données de probabilité d'ouverture d'un e-mail pour un intervalle très court (1 jour) sont à 50%, ce qui est proche du taux moyen global d'ouverture d'un e-mail (54%) sur l'ensemble des agents. En revanche, pour la plupart des intervalles considérés (supérieur à un jour), la probabilité d'ouverture est pour ce groupe beaucoup plus basse (jusqu'à 20%).

Etant donné que 80% des données sur les intervalles entre e-mails se situent entre 7 et 58 jours, on peut recommander sur ce groupe un envoi d'e-mail avec un intervalle très inférieur à 28 jours ou très supérieur à 28 jours, mais pas autour de 28 jours (sinon la probabilité d'ouvrir l'e-mail tombe à 20%).

Sur ce groupe 2, le modèle HLM montre que le nombre de listes de cadeaux et le fait d'avoir donné son numéro de téléphone portable influent sur la probabilité d'ouverture de l'e-mail, comme le montre le tableau 4.19.

4.3.4 Intégration de l'impact de l'e-mail dans le réseau de Petri:

Dans une deuxième étape, nous cherchons à mesurer la performance de campagnes e-mailing en utilisant les réseaux de Petri, en étudiant en particulier l'impact de l'intervalle entre deux e-mails sur les volumes d'achats et sur le nombre de connexions sur le site Millemercis.
L'un des avantages des réseaux de Petri étant de pouvoir mener des simulations (fondées sur des paramètres différents intégrés dans le modèle), nous avons simulé deux types de stratégies d'envoi d'e-mailing en modifiant pour chaque stratégie l'intervalle entre deux e-mails envoyés.
Dans nos données, 80% des intervalles entre 2 e-mails se situent entre 7 et 58 jours (tableau 4.17). En moyenne, Millemercis a envoyé un e-mail publicitaire tous les 23 jours aux agents A.
Nous souhaitons donc comparer la stratégie effectivement menée par Millemercis (un envoi d'e-mail tous les 23 jours) par rapport à une 2ème stratégie qui a priori selon nos analyses précédentes sur les groupes 1 et 2 serait plus performante.
Pour sélectionner l'intervalle d'envoi d'e-mail de cette 2ème stratégie, nous avons analysé plus précisément l'impact de l'intervalle entre deux e-mails, décrit dans le tableau 4.20 ci-dessous. Les deux premières colonnes du tableau présentent la probabilité d'ouverture de l'e-mail pour les groupes 1 et 2 déterminés précédemment. Les deux dernières colonnes montrent le nombre moyen d'e-mails ouverts pendant 60 jours. Elles incluent donc le trade off entre le volume des e-mails envoyés (forte/faible fréquence) et le taux d'ouverture d'un e-mail.

Tableau 4.20 : intervalles entre e-mails, taux d'ouvertures et nombre d'e-mails ouverts par groupe (1 et 2)

	Opening Probability		Number of Opened Emails	
Interval (Day)	Group 1	Group 2	Group 1	Group 2
1	0.160	0.500	9.616	30.000
2	0.292	0.377	8.772	11.315
3	0.371	0.323	7.420	6.460
4	0.421	0.292	6.316	4.383
5	0.455	0.272	5.461	3.268
6	0.479	0.258	4.791	2.585
7	0.497	0.248	4.258	2.129
8	0.510	0.241	3.825	1.805
9	0.520	0.235	3.468	1.564
10	0.528	0.230	3.167	1.379
11	0.534	0.226	2.912	1.233
12	0.538	0.223	2.692	1.114
13	0.542	0.220	2.501	1.017
14	0.545	0.218	2.334	0.935
15	0.546	0.216	2.186	0.866
16	0.548	0.215	2.054	0.806
17	0.549	0.214	1.936	0.755
18	0.549	0.213	1.830	0.709
19	0.549	0.212	1.734	0.670
20	0.549	0.211	1.647	0.634
21	0.549	0.211	1.567	0.602
22	0.548	0.210	1.494	0.574
23	0.547	0.210	1.427	0.548
24	0.546	0.210	1.365	0.525
25	0.545	0.210	1.308	0.503
26	0.544	0.210	1.254	0.484
27	0.542	0.210	1.205	0.466
28	0.541	0.210	1.159	0.450
29	0.539	0.210	1.115	0.434
30	0.537	0.210	1.075	0.420

Le tableau 4.20 montre que pour le groupe 1, le taux d'ouverture atteint son maximum pour un intervalle de 18,9 jours mais le nombre d'e-mails ouverts n'est pas maximisé (nombre d'e-mails ouverts = 1.734 versus 4.258 pour un intervalle de 7 jours). Or il est intéressant dans la pratique de maximiser le taux d'ouverture d'un e-mail mais surtout de maximiser le nombre d'e-mails ouverts sur la période considérée. Il est donc important de tenir compte de ces deux critères pour déterminer la meilleure stratégie marketing à adopter.

4.3.5 Implications managériales :

Parmi les différentes stratégies possibles, une stratégie d'e-mailing tous les 7 jours semble intéressante à simuler pour les raisons suivantes :

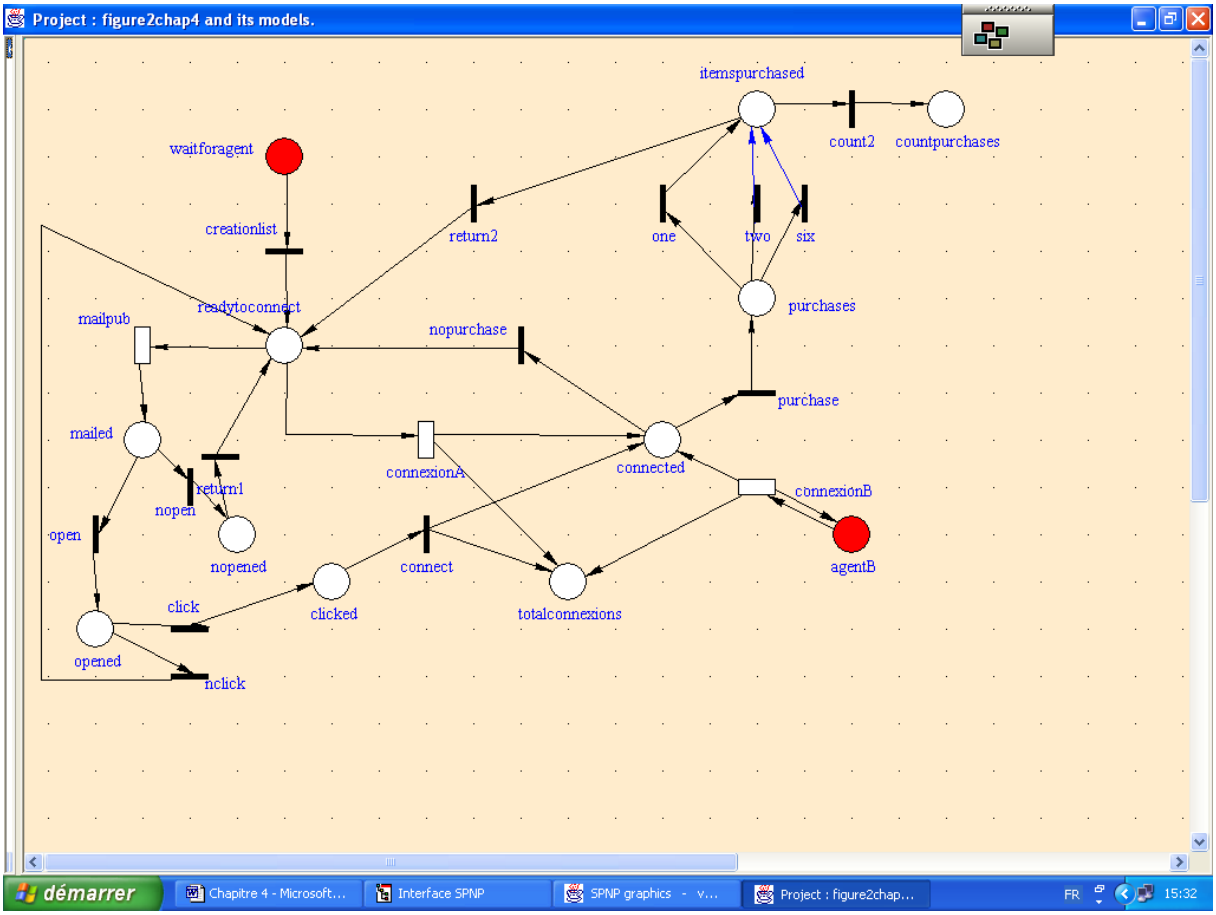
- le taux d'ouverture et le nombre d'e-mails ouverts sont relativement élevés pour chacun des deux groupes (ceci correspond notamment à la probabilité d'ouverture maximale pour le groupe 2 et une probabilité relativement élevée pour le groupe 1 (49.7%) par rapport au maximum obtenu sur ce groupe (54.9%))
- on reste sur un envoi tous les 7 jours, ce qui est « managérialement »envisageable. En effet, il semble peu réaliste de simuler une stratégie d'envoi d'e-mail tous les jours en raison des phénomènes de saturation qui ne sont pas observés de manière suffisante dans les données observées (un envoi tous les jours ne représente que 2,5 % des données observées). De plus l'envoi d'un e-mail par jour n'est pas recommandé par le modèle sur le groupe 1 si l'on tient compte à la fois du critère nombre d'e-mails ouverts et du critère probabilité d'ouverture (tableau 4.20). En effet, pour un intervalle de 1 jour, même si le nombre d'e-mails ouverts est important, la probabilité d'ouverture reste très faible (0.16) versus 7 jours (0.497).

Comparé à l'intervalle de 23 jours (utilisé dans la pratique par Millemercis), l'intervalle de 7 jours peut plus que doubler le nombre d'e-mails ouverts pour le groupe 1 (de 1,43 à 4,25, tableau 4.20) et être multiplié par quatre pour le groupe 2 (de 0,55 à 2,13) pendant 60 jours.
Nous avons donc décidé de simuler ces deux stratégies en utilisant le réseau de Petri : un envoi d'e-mail tous les 7 jours versus un envoi d'e-mail tous les 23 jours. L'objectif est de mesurer l'impact de ces deux stratégies sur le nombre d'items achetés et le nombre de connexions effectuées sur le site Millermercis.

Afin de décrire notre démarche sur les réseaux de Petri, reprenons ici l'intégration de l'e-mailing dans le réseau de Petri, et la figure 4.9 correspondante. L'objectif ici est d'effectuer des simulations grâce au réseau de Pétri afin de comparer deux stratégies marketing : envoi d'un e-mail tous les 7 jours et envoi d'un e-mail tous les 23 jours. L'un des avantages des réseaux de Pétri est de pouvoir en effet modifier les paramètres du modèle proposé en fonction de stratégies marketing possibles et de mesurer ensuite les performances globales du système pour chacune de ces stratégies. Le détail de cette démarche est présentée ci-dessous, suivie des résultats des 2 simulations (7 jours versus 23 jours).

Figure 4.9: login akwell1

Wish list incluant l'e-mailing modélisée par un réseau de Petri



Comme nous l'avons déjà évoqué dans la partie 4.3.2, l'e-mailing est représenté dans le réseau de Petri de la figure 4.9 par la transition *mailpub*. Le taux d'ouverture des emails (probabilité de franchissement de la transition *open*) a été paramétré dans le modèle de Petri avec les équations du modèle HLM (cf partie 4.3.3). Si le mail est ouvert (place *opened*), l'agent A peut cliquer sur cet e-mail (transition *click*) ou ne pas cliquer (transition *noclick*). La probabilité de click paramétrée dans le modèle Petri est la moyenne du groupe auquel appartient l'agent (selon les groupes 1 et 2 déterminés pour le modèle HLM, cf partie 4.3.3). Le reste du modèle de Petri a déjà été présenté.

Nous avons donc élaboré un réseau de Petri incluant l'e-mailing pour chaque agent, en utilisant les paramètres précédemment obtenus pour décrire les connexions des agents A et B sur le site (lois de Weibull) ainsi que la loi NBD pour le comportement d'achat. Puis nous avons simulé l'impact de la fréquence entre deux e-mails sur deux variables : le nombre total de connexions des agents A et B; le nombre d'items achetés. Pour chaque agent, nous avons simulé, via son réseau de Petri sur une période de 16 mois (période maximale commune à un nombre maximum d'agents), deux stratégies d'envoi d'e-mails : un envoi tous les 7 jours, un envoi tous les 23 jours.

Le paramétrage de l'intervalle entre deux e-mails s'effectue dans le réseau de Petri lors de la définition de la transition *mailpub*. Nous avons choisi pour le franchissement de cette transition une distribution constante de paramètres 7 ou 23 jours selon la simulation effectuée.

Les résultats pour chaque agent sont fournis dans le tableau ci-dessous (le croisement des 94 logins comportant 16 mois d'historique et les logins ayant des informations de profils permettant de faire tourner le modèle HLM a donné au final 77 logins pour cette analyse finale) :

Tableau 4,21: résultats de la simulation du modèle Pétri
Comparaison de 2 fréquences d'e-mailing: 7 jours / 23 jours

login	groupe (1 ou 2)	7 jours		23 jours	
		Nbre connexions	Nbre achats	Nbre connexions	Nbre achats
130961	1	353,51	3,37	353,51	3,37
akwell1	2	145,64	45,49	139,93	44,57
AlexO	2	244,35	28,2	235,24	27,19
ama1	2	32,91	0	24,7	0
amibio	2	79,71	0	68,91	0
AnnA	2	121,97	7,94	120,99	7,93
Aude	2	104,13	8,35	95,49	8,03
badir	1	64,76	0	56,73	0
bgougay	2	187,04	2,47	181,77	2,43
bpdl	2	100,33	2,56	91,96	2,36
bruyere	2	48,46	0	42,55	0
bschunke	2	8,67	0,38	8,74	0,36
carolala	2	119,84	0	113,3	0
cathou1	2	45,85	0	36,78	0
chap's	2	64,35	4,79	56,27	4,13
Cypher	2	25,26	4,03	19,49	3,3
DALMATE	1	103,78	0	98,01	0
Delatte	2	126,76	3,07	119,87	2,85
didine77	2	75,34	0	71,44	0
Dinou	2	70,9	6,16	65,59	5,81
djoule	2	187,48	20,16	186,57	20,55
dolores3	2	18,05	0	10,62	0
doppel	2	10,45	0	9,03	0
eb95	2	10,11	0	10,27	0
eb96	2	10,4	0	9,36	0
estia	1	10,95	0	10,99	0
fabou	2	61,72	0	60,19	0
fallouis	1	39,99	3,58	33,33	3,01
fevrard	2	20,73	0,6	13,13	0,39
FOURNIAL	1	40,86	5,22	36,32	4,71
frzpower	1	38,23	0	29,83	0
gallouin	2	619,23	1,63	607,91	1,56
Ginie31	1	386,29	0	381,24	0
Greensky	2	66,85	3,28	58,43	0
gzing	1	59,27	0	53,76	0
hutch	2	99,85	0	93,85	0
jeanbask	1	49,57	0	44,58	0
jeannot3	2	978,34	36,59	969,72	35,61
jeda	1	11,03	0	10,97	0
JIMASA	1	11	0	11	0

Tableau 4.21: résultats de la simulation du modèle Pétri
Comparaison de 2 fréquences d'e-mailing: 7 jours / 23 jours

login	groupe (1 ou 2)	7 jours		23 jours	
		Nbre connexions	Nbre achats	Nbre connexions	Nbre achats
joewj	2	10	0	10,8	0
juboubou	2	97,88	0	87,16	0
kabotine	2	383,3	0	376,25	0
Kifran	2	134,53	0	129,06	0
lbarbier	1	671,87	0	667,03	0
levieux	1	83,91	0	77,38	0
luberdov	2	159,13	0	151,16	0
m@hy	2	176,63	0	161,82	0
mary_ce	2	25,25	0	17,31	0
mekiela	2	119,54	9,83	116,86	9,59
Michel	1	41,68	0	34,91	0
minious	2	45,13	0	40,61	0
mperrot	2	26,14	7,36	24,94	7,27
Naemesis	2	131,3	0	122,35	0
nhai07	2	152,25	0	146,27	0
ninie12	2	10,94	0	10,86	0
philia	1	213,41	1,92	206,92	1,85
picardjf	2	10,93	0	10,81	0
popcorn	1	40,18	0	30,89	0
prioton	2	63,09	0	58,05	0
prozaiik	2	498,4	0	489,37	0
quentin	1	375,29	0	371,93	0
RAPH	2	86,34	0	81,41	0
rlaval	2	110,53	0,84	107,53	0,78
romain37	2	11,01	0	12,23	0
sandor	2	130,09	0	123,49	0
semery	1	134,28	0,86	134	0,85
sophynet	2	120,01	0	112,21	0
Souris	2	10,95	0	11,05	0
starchen	2	7,1	0	2,31	0
stumpf	2	101,36	0	96,46	0
Sylvie4	2	42,96	0	34,09	0
Thibaut	1	457,4	6,08	452,07	6,11
toto2	2	41,83	0	34,23	0
tranb	2	55,4	0	45,52	0
vianney2	2	911,02	2,81	873,7	2,68
Virginie	2	133,5	0	125,75	0

La synthèse du tableau ci-dessus est constitué de la moyenne de chaque variable calculée sur l'ensemble des agents :

	Nbre de connexions	Nbre d'items achetés
Groupe 1		
7 jours	159,36	1,05
23 jours	154,77	1,00
7 jours vs 23 jours	**+ 2,97%**	**+ 5%**
Groupe 2		
7 jours	131,43	3,45
23 jours	125,19	3,29
7 jours vs 23 jours	**+ 5%**	**+ 5%**

Tableau 4.22

Comme on l'avait prévu en comparant le nombre d'e-mails ouverts pour chaque stratégie, l'intervalle entre deux e-mails influence directement la performance en terme de nombre de connexions et volume d'achat.

Tout d'abord, l'amélioration de la performance de l'e-mailing en choisissant la stratégie optimale des 7 jours se constate sur le nombre de connexions. Le nombre de connexions augmente de quasiment 3% pour le groupe 1 et de 5% pour le groupe 2 si l'entreprise opte pour un envoi tous les 7 jours au lieu de sa politique actuelle qui est d'envoyer un e-mail tous les 23 jours. Les deux hausses (groupe 1 et groupe 2) du nombre de connexions sont significatives ($\alpha = 5\%$) puisque nous avons mené un t-test (paired sample) donnant comme p-value 0.00 dans les 2 cas). Ces conclusions sont liées aux hypothèses que nous avons formulées concernant l'appartenance des agents aux deux groupes 1 et 2 tels que définis précédemment.

Par ailleurs, l'amélioration du nombre de connexions entraîne une croissance des volumes d'achat. Pour les deux groupes, le nombre d'items achetés augmente de 5% si l'entreprise implante une nouvelle stratégie d'envoi d'e-mails publicitaires (tous les 7 jours au lieu de 23 jours). La hausse du nombre d'items achetés est également significative dans les deux cas (groupe 1 et groupe 2) puisque nous avons mené un t-test (paired sample) donnant comme p-value 0.15 pour le groupe 1 (résultat significatif pour ce groupe puisque p-value > 0.05) et 0.00 pour le groupe 2.

Ces conclusions sont particulièrement intéressantes au niveau managérial puisqu'elles montrent qu'il est plus performant d'envoyer des e-mails toutes les semaines plutôt que tous les 23 jours.

Cependant, étant donné le nombre important d'e-mails reçus quotidiennement par les internautes, ces résultats sont à modérer en raison de phénomènes de saturation que nous ne mesurons pas dans notre modèle (même si un e-mail tous les 7 jours ne semble pas être a priori très « saturant » pour un internaute).

Ces résultats montrent également la puissance des réseaux de Petri et leur utilisation managériale. En effet, nous avons ici simulé 2 stratégies d'envoi d'e-mail (7 et 23 jours) mais il est possible de simuler de nombreuses autres stratégies (envoi d'e-mail tous les 2 mois, envoi d'e-mail sur les agents B, ou toute autre stratégie marketing menée visant à accroître le nombre de connexions et d'achats des agents A et B sur le site (on peut imaginer un jeu visant à augmenter le traffic par exemple)).

L'intérêt d'utiliser les réseaux de Petri dans le cas qui vient d'être présenté est alors d'aider les responsables marketing dans leur choix entre différentes stratégies possibles, le modèle permettant d'anticiper la performance du système global en fonction des inputs marketing spécifiques à chacune des stratégies testées.

CONCLUSION:

L'objectif de ce chapitre 4 était d'appliquer les réseaux de Petri pour modéliser une première application marketing. Nous avons effectivement modélisé les connexions et le comportement d'achat de différents agents sur le site Millemercis dont le principe repose sur la constitution de listes de cadeaux (cadeaux pouvant être offerts par des amis ou des connaissances). Ce type de site Web constitue un système multi-agents sur Internet, dans lesquels plusieurs agents et plusieurs entreprises cherchent à maximiser leur intérêt spécifique (recevoir un cadeau pour l'agent A, offrir un cadeau pour l'agent B, vendre une produit pour l'entreprise).
Nous avons en particulier montré qu'une modélisation par les réseaux de Petri au niveau individuel donnait de meilleurs résultats sur le nombre d'items achetés qu'un modèle NBD avec régression (sur des variables socio-démographiques de l'internaute), ce dernier type de modèle étant fréquemment utilisé en marketing. En utilisant une modélisation par agent dans le modèle de Petri, nous avons également soulevé une limite de notre approche qui est le nombre important de paramètres du modèle proposé. Dans le cas que nous modélisons dans les chapitres 5 et 6 de la thèse, nous utilisons un modèle sur l'ensemble des individus qui démontre également la flexibilité du modèle de Petri qui peut être mené au niveau individuel ou global.
Dans la constitution du modèle de Petri, et contrairement à des recherches précédentes, nous incorporons différentes actions dans une structure flexible et cherchons à expliquer l'impact de ces actions sur le comportement d'achat. Nous montrons que les réseaux de Petri (stochastiques dans notre cas) permettent de décomposer un système complexe en différents blocs de « sous-modèles » qui chacun décrivent un phénomène spécifique (comportement de connexion des agents A, comportement de connexion des agents B, comportement d'achat des agents A et B) en utilisant la théorie des probabilités et des modèles stochastiques. Nous relions ensuite ces « sous-modèles » de manière dynamique pour représenter le système global du fonctionnement des wish lists. Ensuite, nous montrons qu'un autre avantage des réseaux de Petri est la possibilité de mener plusieurs simulations de stratégies différentes. Nous simulons dans ce chapitre deux stratégies d'envoi d'e-mails (tous les 7 jours, tous les 23 jours) et déterminons l'impact de chacune de ces stratégies sur la performance globale du système (définie ici par 2 variables : nombre de connexions, nombre d'items achetés). Nous montrons que la stratégie d'envoi d'e-mail tous les 7 jours est plus performante que celle qui

consiste à envoyer un e-mail tous les 23 jours. Plus globalement, notre modèle enrichit des approches comme celles de Sismeiro et Bucklin (2004) qui analysent le processus de visite et d'achat de manière séquentielle, le réseau de Petri représentant un système de wish list qui intègre à la fois des agents A créant leur liste et des agents B se connectant en parallèle pour consulter la wish list des agents A. Ce type d'action parallèle ne peut pas être analysée par des modèles uniquement séquentiels. L'un des avantages d'utiliser les réseaux de Petri est donc de pouvoir intégrer des actions parallèles, synchronisées et même des systèmes de chaînes de Markov non stationnaires. Nous montrons ainsi que les réseaux de Petri stochastiques sont particulièrement bien adaptés pour modéliser des phénomènes sur Internet dans lesquels les systèmes sont souvent complexes, comportant des processus stochastiques, des actions parallèles ou synchronisées de différents agents, avec de plus un nombre de données souvent très important (nous avons travaillé ici sur plus de 20 000 connexions). Les réseaux de Petri simplifient cette complexité et fournissent une représentation graphique simple du phénomène étudié, permettant à la fois aux chercheurs mais aussi aux managers de mieux comprendre le système étudié et sa modélisation.

Cependant, l'objectif étant de présenter les réseaux de Petri sur une première application « simple » pour des raisons de compréhension de la méthode utilisée, notre modèle dans ce chapitre n'inclut pas tous les aspects liés au comportement d'achat sur le Web. Une première limite est la distribution de probabilité utilisée pour mesurer le nombre d'items achetés. Dans notre modèle, nous avons appliqué un modèle NBD sans tenir compte du nombre d'items présents sur la liste, ni du type d'items (catégorie de produit, prix). On aurait pu utiliser par exemple une distribution binomiale dans laquelle l'information sur le nombre exact d'items de la liste de cadeaux serait fourni, ou intégrer plus de critères liés au type de listes de cadeaux constituées. Nous avons également utilisé dans ce chapitre une modélisation par individu avec 6 paramètres à calculer par individu, ce qui rend le modèle plus complexe mais montre également que les réseaux de Pétri peuvent être utilisé pour des modèles par individu. Cette limite sera levée dans les chapitres 5 et 6 dans lesquels nous présenterons une deuxième application dans laquelle nous travaillerons sur l'ensemble des individus. Une autre limitation dans ce chapitre est que nous avons sélectionné un échantillon de listes actives sur lesquelles nous avons travaillé. Nous développerons une application dans les prochains chapitres utilisant l'ensemble des données disponibles. On pourrait également complexifier le réseau en intégrant la navigation de l'internaute sur le site (pages vues, durée de visite, etc…), ce qui est tout à fait possible de par le principe du modèle de Petri. Une autre limitation concerne les stratégies d'e-mailing testées. D'une part, nous ne mesurons pas dans notre modèle les effets

de saturation liés à des e-mails répétés. D'autre part, notre modèle a été appliqué sur les campagnes e-mailing effectuées par une seule entreprise dans le contexte particulier des listes de cadeaux, et le choix des stratégies testées de 7 jours et de 23 jours est donc lié aux actions menées par cette entreprise. Il serait intéressant d'appliquer le modèle à d'autres sites Web et d'autres stratégies d'envoi de campagnes e-mailing.

Les limites que nous avons soulevées dans cette application nous ont poussé à appliquer les réseaux de Petri à une deuxième application marketing sur Internet qui comporte moins de spécificités que les listes de cadeaux et qui permet de mener des simulations sur plusieurs inputs marketing. Nous avons donc choisi de modéliser une opération de marketing viral dans les chapitres 5 et 6 de la thèse.

Chapitre 5
Modélisation d'un jeu viral promotionnel on line par les réseaux de Petri

5.1 INTRODUCTION :

L'objectif de ce chapitre est d'appliquer la méthodologie des réseaux de Petri à un deuxième cas dans le domaine du marketing électronique. Nous avons choisi d'étudier plus particulièrement l'impact de campagnes de marketing viral sur Internet. En effet, le marketing viral, ou bouche-à-oreille via le Web, est une technique utilisée par les entreprises en raison de son faible coût, de sa facilité d'opérationnalisation et de son efficacité reconnue. L'exemple d'Hotmail illustre ces trois points : Hotmail est un service d'e-mail gratuit qui a développé le nombre de ses clients de la manière suivante : la ligne « Get Your private, Free Email at http://www.hotmail.com » était incluse dans tout e-mail envoyé depuis un compte Hotmail. Lancé en Juillet 1996, 12 millions d'utilisateurs ont signé en deux ans, pour un budget marketing de $500 000 (Stambouli & Briones, 2002).

Etudier l'impact d'actions de marketing viral répond à un besoin actuel des entreprises. En effet, les managers aujourd'hui s'interrogent sur ces techniques. Tout d'abord, quel mécanisme mettre en place ? Quelles actions mener ? En effet, comme il n'existe pas aujourd'hui de « technique homologuée » en terme de marketing viral, les entreprises sont dans une phase d'essai de différents types de campagnes et acquièrent une compétence liée à leur expérience. D'autre part, les entreprises souhaitent mesurer l'efficacité de ce type d'action, mais il n'existe pas de modèles reconnus sur le marché pour mesurer l'impact de campagnes de marketing viral. Nous proposons donc dans ce chapitre d'introduire les réseaux de Petri pour modéliser ce type d'action et en mesurer l'efficacité. Ceci repose sur le fait que le phénomène viral va se propager au sein d'un réseau d'individus qui vont être reliés entre eux par des liens ou des actions, l'ensemble du système pouvant être représenté par un réseau de Petri.

Notre cas d'application (l'opération « Le ciel est à tout le monde »), est un jeu promotionnel on line visant à créer une base de données qualifiée d'adresses e-mail. Par rapport au thème général de la thèse, ce jeu constitue bien un système multi-agents on line dans lequel plusieurs agents internautes vont interagir avec l'interface Web du site pour tenter de gagner. Ce système peut devenir rapidement complexe à étudier et à modéliser de par le nombre

important de joueurs et la mécanique plus ou moins complexe du jeu utilisé. Les réseaux de Petri sont une réponse à ces problématiques.

Ce chapitre est structuré de la manière suivante : dans une première partie sont présentés plusieurs articles portant sur le marketing viral ; la deuxième partie décrit le mécanisme du jeu promotionnel « Le Ciel est à tout le monde » et la stratégie de l'entreprise associée (Anyway) ; la troisième partie porte sur l'application des réseaux de Petri pour modéliser ce jeu on line. Les résultats constituent la quatrième partie.

Nous concluons ensuite sur l'intérêt d'utiliser les réseaux de Petri pour modéliser les systèmes multi-agents interactifs que sont les jeux promotionnels on-line.

5.2 REVUE DE LITTERATURE SUR LE MARKETING VIRAL SUR LE WEB :

La littérature en marketing est peu fournie sur le thème du marketing viral. Elle est en revanche riche d'articles concernant les phénomènes de bouche-à-oreille. Certains articles, issus de la littérature managériale, développent même l'idée selon laquelle les processus de communication fondés sur le bouche-à-oreille constituent une force puissante sur un marché (Marney, 1995 ; Silverman, 1997). Le bouche-à-oreille (BAO) correspond à une communication informelle, à la fois positive et négative, sur les caractéristiques d'un fournisseur et/ou ses produits (Tax et al. 1993). Utilisant le même mécanisme que le bouche-à-oreille mais sur Internet, la notion de marketing viral a été initialement utilisée par Steve Jurvetson en 1996 pour décrire la stratégie marketing du service gratuit d'e-mail Hotmail. Le marketing viral a fait l'objet de nombreux articles managériaux mais de très peu d'articles académiques. Les articles managériaux se résument essentiellement à la présentation de cas d'entreprises à succès. Les exemples comme Beanie Babies, the Blair Witch Project, Paypal, dont la stratégie repose sur le phénomène de marketing viral sont très fréquemment cités. Business Week (2001) annonce même le nouveau monde du buzz marketing, citant plusieurs exemples menés durant l'été 2001, comme les stratégies de Vespa, des jeux Hasbro ou encore de Ford. Nous nous concentrerons par conséquent sur les nombreuses recherches académiques concernant le phénomène du bouche-à-oreille (base du marketing viral), ce thème n'étant en revanche pas récent dans la littérature en marketing. En 1955, Katz et Lazarsfeld mettent en avant que le BAO serait sept fois plus efficace qu'une publicité dans un journal ou un magazine, quatre fois plus efficace qu'une vente personnelle, deux fois plus efficace qu'une campagne radio pour influencer le consommateur à changer de marque. Arndt (1968) montre

que les individus les plus disposés à acheter un nouveau produit étaient aussi ceux les plus exposés au bouche-à-oreille. Les individus ayant été exposés au bouche-à-oreille sont également plus favorablement prédisposés à l'achat. Arndt (1967b) a également mené une revue de littérature sur le bouche-à-oreille en communication. Il montre notamment l'effet des conversations sur un produit dans la diffusion d'un nouveau produit (Arndt (1967a)).

Nous retiendrons dans cette littérature sur le BAO deux recherches en particulier qui proposent d'analyser le bouche-à-oreille par des analyses de réseau, notre modèle (les réseaux de Petri) étant fondé sur la même logique. Le premier article de Brown et Reigen (1987) est un article de référence sur le BAO présentant une analyse de réseau du comportement de bouche-à-oreille dans un environnement naturel. Les auteurs examinent le comportement de bouche-à-oreille grâce à la force de lien et à l'homophilie (degré auquel des paires d'individus sont similaires sur certains attributs comme l'âge, le sexe et l'éducation). Leur étude met en évidence au niveau macro et micro les différents rôles joués par les liens sociaux faibles et forts. Au niveau macro, les faibles liens exercent une fonction de passerelle, permettant à l'information de circuler d'un sous-groupe d'acteurs à un autre sous-groupe distinct dans un système social plus large. Au niveau micro, les forts liens et l'homophilie sont activés pour développer le flux d'informations auquel les individus vont se référer. Le deuxième article, de Bristor (1990), propose aussi une approche de réseau pour étudier le bouche-à-oreille. Selon cet auteur, l'approche par réseau est une alternative prometteuse pour l'étude du bouche-à-oreille car elle met en évidence les relations qui lient les membres d'un système social. Bristor affirme qu'un réseau de bouche-à-oreille est un réseau social de personnes qui s'engagent les unes envers les autres au delà des relations classiques qui les relient. Certains auteurs (Bansal et Voyer, 2000) ont analysé le processus de bouche-à-oreille dans un contexte de décision d'achat de service. Leur recherche met en évidence trois relations distinctes: l'effet de forces non interpersonnelles (l'expertise du receveur, le risque perçu du receveur, l'expertise de l'expéditeur) ; l'effet de forces interpersonnelles (force des liens, manière dont le bouche-à-oreille est recherché) ; les effets des forces non interpersonnelles sur les forces interpersonnelles. D'autres auteurs se sont intéressés aux antécédents du bouche-à-oreille, comme Bowman et Narayandas (2001), qui étudient l'impact de contacts avec le consommateur, qu'ils nomment CIC (customer-initiated contacts) sur le comportement de bouche-à-oreille. Ils montrent notamment que lorsque les consommateurs s'engagent dans un comportement de bouche-à-oreille suite à un contact CIC, le nombre médian de consommateurs influencés est approximativement de 3, quelles que soient les caractéristiques du consommateur. Harrison-Walker (2001) développe une échelle de mesure de la

communication par bouche-à-oreille et identifie deux variables comme antécédent du bouche-à-oreille : l'engagement du consommateur et la qualité du service. L'engagement affectif est relié positivement au bouche-à-oreille tandis que l'engagement comme sacrifice n'est pas relié à ce type de communication. Quant à la qualité de service, l'auteur montre que ses effets sur le bouche-à-oreille dépendent du type d'industrie.

D'autres articles portent plus précisément sur le concept de marketing viral sur le Web. Récemment, Kaikati et Kaikati (2004) développent le concept de marketing furtif ancré sur le principe selon lequel le bouche-à-oreille reste la forme de promotion la plus efficace. Selon ces auteurs, les individus sont aujourd'hui bombardés de messages publicitaires et développent en conséquence des mécanismes de défense. Les marketers doivent donc utiliser des méthodes plus subtiles pour communiquer un message aux consommateurs, comme le marketing furtif qui consiste à présenter un nouveau produit en créant intelligemment un « buzz » de manière clandestine. Plusieurs techniques marketing sont proposées par les auteurs dans le cadre du marketing furtif : le marketing viral, les « pousseurs de marque » (jeunes acteurs s'approchant de personnes dans des situations de vie réelle en leur glissant des messages commerciaux dans des bars branchés, des magasins de musique), le marketing célébrité (des célébrités sont embauchées par des laboratoires pharmaceutiques pour discuter et mentionner certains médicaments), le marketing dans les jeux vidéo (faire apparaître les marques ou les logos pour les renforcer) et enfin le marketing dans la musique Rap et Pop. Le marketing viral a été analysé sous l'angle du consommateur-vendeur par Helm (2000) qui souligne l'idée selon laquelle de nombreuses entreprises essaient d'utiliser sur le Web les réseaux de communication de leurs propres consommateurs pour promouvoir et distribuer leurs produits. L'expansion d'Internet ouvre également de nouvelles perspectives en ce qui concerne le bouche-à-oreille. Certains sites (Walmart, Chase Manhattan Bank, Toys 'R' Us) proposent des espaces de discussion sur les produits et les services multipliant ainsi les effets du bouche-à-oreille. L'entreprise a donc intérêt à utiliser ces interactions entre internautes. Helm (2000) définit le marketing viral comme un concept de communication et de distribution qui repose sur des consommateurs transmettant des produits digitaux par e-mail à d'autres consommateurs potentiels de leur sphère sociale et animant ces contacts pour que ceux-ci transmettent aussi les produits à d'autres. L'auteur décrit les trois caractéristiques d'une stratégie de marketing viral efficace : le produit est proposé gratuitement uniquement sur le Net et pas dans les autres circuits de distribution (les très faibles coûts fixes sur Internet favorisent la distribution gratuite des produits, le coût marginal pour une unité supplémentaire est proche de zéro) ; l'offre contient une réelle valeur ajoutée ; les premiers contacts sont

choisis avec précaution. Helm (2000) distingue également les stratégies de marketing viral avec forte ou faible intégration du consommateur. Dans un article provocateur, Dye (2000) montre que le marketing viral n'obéit pas aux cinq mythes suivants pourtant très répandus : les produits à scandale ne sont pas les seuls produits intéressants à promouvoir par bouche-à-oreille (certains médicaments peuvent créer un bouche-à-oreille très fort) ; le buzz n'arrive pas tout seul mais est le résultat d'actions marketing bien organisées ; les meilleurs déclencheurs de buzz ne sont pas les meilleurs consommateurs (certains sont à contre courant de la société) ; pour profiter de l'effet d'un buzz, on ne doit pas agir en premier et vite (certains « copieurs » peuvent tirer des profits substantiels s'ils savent quand créer le buzz) ; les media et de la publicité ne sont pas nécessaire pour créer un buzz.

Dans un ouvrage consacré au marketing viral, Seth Godin (2001) souligne que le marketing viral est lié au fait que les individus sont plus connectés que jamais entre eux. Pour affirmer ce propos, il compare le faible nombre de personnes auxquelles un individu téléphonait il y a dix ans environ (une dizaine, voire une trentaine dans une vie personnelle) par rapport à la liste de contacts d'une messagerie vocale et le nombre de personnes par semaine avec lesquelles un internaute échange des e-mails aujourd'hui. Dans cet ouvrage, Godin (2001) définit huit variables fondamentales définissant le marketing viral. La première est le *contaminateur*, défini comme une personne prédisposée à parler autour d'elle d'une nouvelle idée. L'auteur distingue deux catégories de contaminateurs : les efficaces, souvent des leaders d'opinion, comme William Shatner par exemple, autrefois légendaire pour son rôle dans *Star Strek*, qui chante les louanges de Priceline, société permettant d'acheter des billets d'avion aux enchères; les erratiques qui essaient en toute circonstance de « vendre » leur idée virus à n'importe qui, le contaminateur erratique pouvant agir dans la perspective de gagner de l'argent : le site Amazon.com par exemple offre des incitations financières aux utilisateurs qui diffusent des idées autour d'eux ; le site Referrals.com, entreprise fondée sur l'idée de rémunérer ceux qui aident les autres à trouver un emploi, demande à des internautes de transmettre l'offre d'emploi à leurs amis. Si l'un d'eux l'accepte, l'internaute contaminateur reçoit un chèque de 4000 euros (cette stratégie se substitue à celle plus classique de confier à un chasseur de tête de trouver, pour 30 000 Euros, 5 candidats correspondant au profil recherché). Selon Godin (2001), le triomphe d'une idée virus commence par le choix de la *ruche* (deuxième variable). Le principe de la ruche est de choisir un marché qui se distingue par une bonne concentration de contaminateurs, comme l'a fait Ebay par exemple. La troisième variable importante en marketing viral selon Godin (2001) est la *vélocité*, rapidité avec laquelle l'idée va se répandre sur Internet.Godin cite comme exemple de vélocité la

société Napster permettant aux internautes de partager les fichiers MP3, concept qui s'est propagé très rapidement dans le monde entier via Internet. La quatrième variable est le *vecteur* que suit toute idée virus : une idée ne se diffuse jamais de façon régulière et harmonieuse dans une population, les individus ayant au contraire tendance à l'envoyer dans tel sens plutôt que tel autre. Le vecteur d'une idée virale peut être par exemple Internet ou des cartes à échanger entre des enfants. La cinquième variable introduite par Godin (2001) pour définir le marketing viral est le *véhicule*, qui explique par exemple le succès de PayPal.com (paiement sécurisé peer to peer sur Internet) : pour chaque achat sur Ebay ou d'autres sites de ventes aux enchères, lors du règlement de l'achat, l'internaute doit virer la somme demandée à PayPal.com, le véhicule dans ce cas étant par conséquent le système de paiement. La *facilité* est la sixième variable évoquée par Godin (2001) : le but du marketing viral est de déclencher un virus offrant une telle facilité que l'individu devient « accro » dès le premier contact, comme le cas de Vindingo avec ses Palm Pilot). La septième variable est la *persistance* de l'idée virus, afin que l'idée virus ne soit pas sans lendemain. Enfin, Godin (2001) évoque une dernière variable : l'*amplificateur*: selon l'auteur, le bouche-à-oreille ne suffit pas et il est nécessaire de mettre en place un système favorisant l'amplification du bouche-à-oreille positif. L'auteur cite à ce propos l'exemple de la société Planetfeedback.com : tout consommateur mécontent ou content d'une entreprise peut exprimer son opinion sur Planetfeedback.com qui la transmettra à celle-ci par courrier électronique. Ensuite, un seul clic permet de faire envoyer une copie de l'e-mail aux administrations, aux medias et aux élus politiques concernés. Un autre clic permet d'envoyer une copie à des amis ou collègues les plus proches du consommateur, l'ensemble de ce système créant donc une amplification instantanée.

Tableau 5.1: Douze idées virus qui méritent réflexion. Source: S.Godin (2001), le marketing viral

Entreprise	Idée choc	Véhicule de diffusion
Polaroïd	Photographie instantanée	"Hé! Regardez donc ça!" vous exclamez-vous lors d'une soirée
Tupperware	Le meilleur système de rangement pour aliments	Arrangez-vous pour que vos amis vendent à leurs amis
Fax	Documents livrés par téléphone	Plus vous vendez à vos associés en affaires, et mieux votre machine fonctionne

A titre d'exemple, douze idées virus soulignées par S. Godin sont synthétisées ci-dessous :

Home Shopping Network	Les achats via la télévision par câble	"Hé! Regardez ce qu'ils vendent sur Home Shopping", apprenez-vous à vos amis par téléphone
Fast Company	Journal de la Nouvelle Economie	Company of Friends: réunion mensuelle des lecteurs locaux du magasine
Camine's Restaurant	Des tonnes de nourriture, des tonnes d'ail	Réservations à partir de six personnes Vous devez vendre l'idée à vos amis avant d'entrer
Beanie Babies	Peluches de collection	Si d'autres personnes commencent à en accumuler, votre collection prend de la valeur
Gamesville	Des jeux qui vous scotchent sur le Web	Grâce à la "souris-à-oreille", envoyez un e-mail à vos amis pour les inviter
Hotmail	Courrier électronique gratuit	Complètement viral; Chaque mail envoyé en fait de la publicité
Tommy Hilfiger	Vêtements dze style BCBG tirant sur le chic	Viru du logo: plus vous le portez, plus les gens le voient
"La cathédrale et le bazar"	Logiciel ouvert et gratuit	Permet au contaminateurs efficaces de répandre ue idée en leur donnant un manifeste fort à partager
Vindigo	Guide touristique Zagat sur le Palm	Un ami peut l'importer en quelques secondes

Cette synthèse de la littérature sur le marketing viral montre par conséquent que peu de recherches académiques portent sur ce thème et en particulier sur la modélisation des opérations de marketing viral. Plus particulièrement, les managers manquent aujourd'hui de modèles les aidant à mieux designer leur campagne de marketing viral, à mieux mesurer les performances de ces opérations. Plus globalement, que ce soit au niveau académique ou appliqué, des recherches sont à mener pour mieux comprendre les phénomènes viraux sur Internet. L'objectif de cette recherche est donc de proposer un modèle permettant de mesurer et d'optimiser l'impact d'actions de marketing viral sur Internet.

A partir de données issues d'une campagne de marketing viral menée par le site de voyages Anyway .com (jeu promotionnel « Le ciel est à tout le monde »), nous appliquerons la

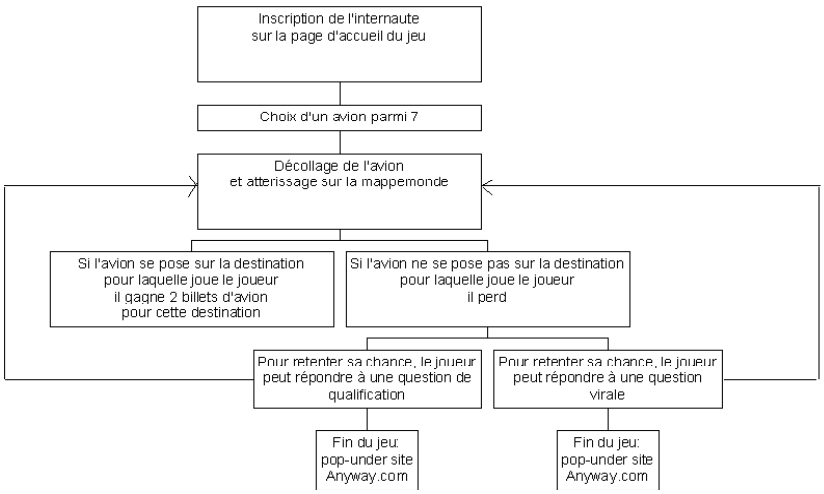
méthode des réseaux de Petri et montrerons que ce modèle est particulièrement adapté à la mesure de l'efficacité « virale » de cette opération marketing.

5.3 PRINCIPE DE L'OPERATION « LE CIEL EST A TOUT LE MONDE » :

L'opération « Le Ciel est à tout le monde » est un jeu promotionnel viral élaboré par la société Millemercis.com pour le compte de la société Anyway. Millemercis.com est le spécialiste français de génération de bases de données et de marketing viral sur Internet. Anyway.com appartient au Groupe IAC (Nasdaq : IACI), dont l'objectif est de devenir l'entreprise de commerce interactif la plus importante au monde grâce à une stratégie multi-marques. Depuis 15 ans, Anyway commercialise des voyages via le site Anyway.com qui intègre dans sa base de données plus de 1 500 000 tarifs négociés. Les produits proposés sur le site sont des vols internationaux et domestiques, des séjours et des circuits, des idées de week-end, des séjours à thèmes, des réservations hôtelières et des locations de voiture. L'opération de marketing viral « Le Ciel est à tout le monde » s'est déroulée entre le 10 Mai et le 8 Juin 2004 (soit une durée de 30 jours) avec comme principal objectif de générer une base de données de prospects et clients qualifiés. Le nombre visé de participants à l'opération était de 200 000. « Le Ciel est à tout le monde » visait également à accompagner le repositionnement d'Anyway : recentrage vers le vol sec, installation du claim « Le Ciel est à tout le monde », visibilité de la nouvelle charte. Plus globalement l'objectif d'Anyway était de relancer le trafic et les ventes.

La mécanique du jeu promotionnel utilisé est décrite dans la figure 5.1 suivante :

Figure 5.1: Mécanisme du jeu

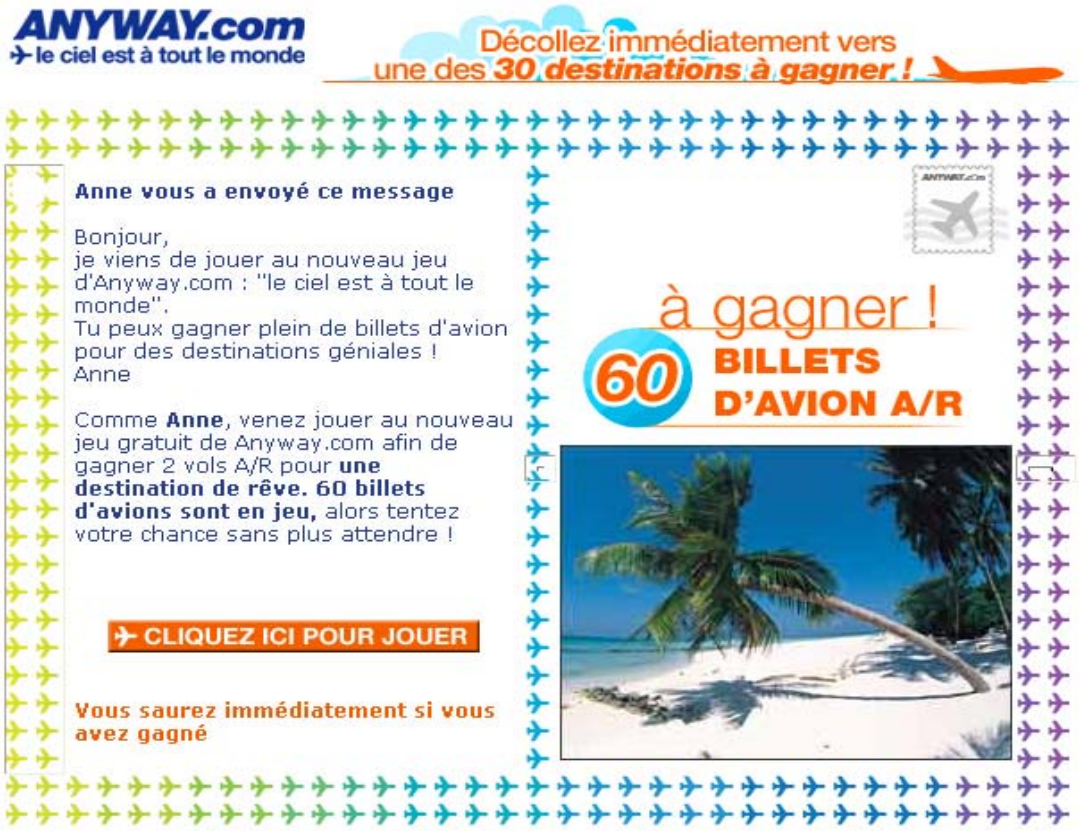


1. L'internaute s'inscrit sur la page d'accueil du jeu,
2. Il choisit ensuite son avion parmi 7,
3. Son avion s'envole et atterrit sur la mappemonde,
4. S'il se pose sur la destination pour laquelle joue le joueur, il a gagné 2 billets d'avion pour cette destination,
5. Sinon, il a perdu. Pour retenter sa chance, le participant doit répondre à une question soit de qualification (socio-démographique ou comportement de voyage), soit virale (donner l'adresse e-mail d'un ami pour lui envoyer une e-card et lui proposer de découvrir le jeu)
6. Lorsque le participant souhaite arrêter de jouer, un pop-under met en avant le site Anyway.com.

Ce mécanisme repose donc sur des actions séquentielles de l'internaute. Celui-ci, pour rejouer plusieurs fois, va répondre à des questions et par conséquent alimenter la base de données qualifiées lorsqu'il répond à une question de qualification, ou augmenter le nombre d'adresses e-mail de la base de données lorsqu'il répond à une question virale.

Les dotations mises en jeu correspondent à 30 billets d'avion aller retour pour deux personnes, soit un gagnant par jour.

A titre d'exemple sont présentés ci-dessous un exemple de question virale ainsi que l'e-card envoyée à l'adresse e-mail virale obtenue.



L'opération « Le ciel est à tout le monde » a été un succès.

Les principaux résultats sont les suivants : 561 535 visites sur 30 jours de jeu, 227 879 joueurs inscrits (pour un objectif initial de 200 000), une durée moyenne de visite de 7 minutes 48 secondes, 10 834 075 pages vues sur la durée du jeu, 472 757 adresses e-mail saisies par les participants.

Le questionnaire complet comporte 41 questions (virales ou de qualification).

L'ensemble des données récoltées via ce questionnaire comporte plus de 2 millions de réponses aux questions, avec une moyenne de 9,27 questions répondues par participant. Plus du quart des participants ont répondu à plus de 15 questions.

3 163 652 données ont ainsi été recueillies durant l'opération « Le ciel est à tout le monde » . 86% de ces données sont des données de qualification (conformément au brief d'Anyway) et 14% des données de viral. Les données de qualification portent essentiellement sur les critères socio-démographiques et les comportements de voyage, les données de viral étant uniquement des adresses e-mail fournies. Le détail du type de données de qualification et des questions virales est fourni dans le questionnaire du jeu présenté ci-dessous :

Tableau 5.2 : Questionnaire du jeu

Q1		Vous vivez	En couple sans enfant	Seul(e) sans enfant	En couple avec enfant(s)	Seul(e) avec enfant(s)	Chez vos parents						
101			1	2	3	4	5						
Q2		Actuellement, vous êtes	Etudiant(e), lycéen(ne), collégien(ne)	Salarié(e) secteur privé	Travailleur(se) indépendant(e)	Fonctionnaire	En recherche d'emploi	H/F au foyer	Retraité(e)	Autre			
201			1	2	3	4	5	6	7	8			
		Merci de préciser si vous êtes…	Agriculteur	Artisan	Commerçant	Chef d'entreprise	Cadre	Cadre supérieur	Enseignant	Profession libérale	Employé	Ouvrier	Autre
202			1	2	3	4	5	6	7	8	9	10	11
Q3		Combien de voyage(s) aller-retour avez-vous effectué(s) au cours des 12 derniers mois ?	Aucun	1	2	3	4	5	6 à 10	11 à 20	21 et plus		
301		En avion pour des raisons personnelles	1	2	3	4	5	6	7	8	9		
302		En avion pour des raisons professionnelles	1	2	3	4	5	6	7	8	9		
Q4		VIRAL											
Q5		En général, pour vos voyages en avion, vous réservez…	Moins de 3 jours à l'avance	Entre 3 et 6 jours à l'avance	Entre une semaine et un mois à l'avance	Entre un mois et 2 mois à l'avance	Plus de 2 mois à l'avance	Je ne réserve pas de billets d'avion					
501			1	2	3	4	5	6					
Q6		Quel est l'aéroport de départ que vous utilisez le plus fréquemment ?											
601		menu déroulant, voir onglet correspondant											
Q7		Quand prévoyez-vous de partir en voyage en avion ?	En mai ou juin 2004	En juillet 2004	En août 2004	En septembre 2004 ou plus tard	Je ne prévois pas de partir en voyage en avion						
701			1	2	3	4	5						
		Où souhaitez vous partir pour votre prochain voyage en avion ?	En France métropolitaine	En Europe du Sud	Dans le reste de l'Europe	Au Moyen Orient	Aux Caraïbes	Dans l'Océan Indien	En Amérique du Nord	En Amérique latine	En Afrique du Nord	Dans le reste de l'Afrique	En Asie
702			1	2	3	4	5	6	7	8	9	10	11
Q8		VIRAL											
Q9		Quelle part de vos voyages achetez-vous sur Internet ?	Je n'en ai jamais acheté sur Internet	Moins d'un quart	Entre un quart et la moitié	Entre la moitié et les trois quarts	Plus des trois quarts						
901		Pour des billets d'avion personnels	1	2	3	4	5						
902		Pour des billets d'avion professionnels	1	2	3	4	5						
Q10		Au cours des 12 derniers mois, combien de séjour(s) en hôtel avez-vous effectué ?	Aucun	1	2	3	4	5	6 et plus				
1101		pour des raisons personnelles	1	2	3	4	5	6	7				
1102		pour des raisons professionnelles	1	2	3	4	5	6	7				

Question	Condition	Intitulé des questions	modalité réponse									
Q11		VIRAL										
Q12		Combien de fois avez-vous loué une voiture au cours des 12 derniers mois ?	Aucun	1	2	3	4	5	6 à 10	11 à 20	21 et plus	
1201		Pour des raisons personnelles	1	2	3	4	5	6	7	8	9	
1202		Pour des raisons professionnelles	1	2	3	4	5	6	7	8	9	
Q13		Pourquoi quelle(s) raison(s) n'achetez vous pas vos voyages personnels sur Internet ? *(Plusieurs réponses possible)*										
1301	Q901=1	Achat en agence de voyage										
1302	Q901=1	Prix trop chers										
1303	Q901=1	Délais de livraison trop importants										
1304	Q901=1	Paiements pas suffisament sécurisés										
1305	Q901=1	Données non confidentielles										
1306	Q901=1	Plaintes ou réclamations difficiles										
1307	Q901=1	Problème de connexion Internet										
1308	Q901=1	Autres										
1309	Q901=1	Je n'achète pas personnellement de voyage										
1399	Q901<>1	VIRAL										
Q14		Au cours des 12 derniers mois, combien de séjours à l'étranger avez-vous effectué pour des raisons personnelles	Je n'ai effect	1	2	3	entre 4 et 6	plus de 6				
1401		Des séjours courts (1 à 4 jours)	1	2	3	4	5	6				
1402		Des séjours d'une semaine	1	2	3	4	5	6				
1403		Des séjours de plus d'une semaine	1	2	3	4	5	6				
Q15		Où êtes-vous parti ces 12 derniers mois ?	En France	En Europe du	Dans le reste	Au Moyen	Aux Caraibes	Dans l'Océan	En Amérique	En Amérique	En Afriq ue du	Dans le reste	En Asie	En Océanie	Je ne sais pas
1501	Q301<>1	pour des raisons personnelles	1	2	3	4	5	6	7	8	9	10	11	12	13
1502	Q302<>1	pour des raisons professionnelles	1	2	3	4	5	6	7	8	9	10	11	12	13
1599	Q301=1 ET	VIRAL													
Q16		VIRAL													
Q17		Quelles sont les 3 compagnies aériennes que vous utilisez le plus souvent ?	Air Fr	Alitali	Amer	Britis	Britis	Delta	Easy	Iberia	KLM	Lufth	Ryan	SAS	Swiss	Autres Compagni
1701	01<>1 ou Q302<	choix 1	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1702	01<>1 ou Q302<	membre du programme de fidélité	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1703	01<>1 ou Q302<	choix 2	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1704	01<>1 ou Q302<	membre du programme de fidélité	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1705	01<>1 ou Q302<	choix 3	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1706	01<>1 ou Q302<	membre du programme de fidélité	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1799	Q301=1 ET	VIRAL														
Q18		Combien d'enfants avez-vous ?	Aucu	1	2	3	4	5								
1801	Q101 IN (2,4)	Moins de 3 ans	1	2	3	4	5	6								
1802	Q101 IN (2,4)	Entre 3 et 7 ans	1	2	3	4	5	6								
1803	Q101 IN (2,4)	Entre 7 et 12 ans	1	2	3	4	5	6								
1804	Q101 IN (2,4)	Entre 12 et 18 ans	1	2	3	4	5	6								
1805	Q101 IN (2,4)	Plus de 18 ans	1	2	3	4	5	6								
1899	Q101<>(2,4)	VIRAL														
Q19		Quels sont les principaux critères qui interviennent dans votre choix pour acheter un billet d'avion (classement par ordre d'importance)	Le prix le	Vol direct	Horaires	Les services	La renommé	Vol en prom	L'existence d'un							
1901		choix 1	1	2	3	4	5	6	7							
1902		choix 2	1	2	3	4	5	6	7							
1903		choix 3	1	2	3	4	5	6	7							
Q20		VIRAL														
Q21		Lorsque vous partez en vacances, vous préférez ...	Un séjour	Un séjour en	Un séjour en	Réserver un	Une croisière	Une thalassoth	Un séjour	Partir à l'ave	Aller dans votre	Aller chez des	Cela dépend			
2101			1	2	3	4	5	6	7	8	9	10	11			
Q22		Lorsque vous voyagez en avion, vous voyagez principalement...	En premi	En class	En class	Je ne voya										
2201		à titre personnel	1	2	3	4										
2202		à titre professionnel	1	2	3	4										
Q23	Q302<>1	Qui s'occupe habituellement de réserver vos billets d'avion professionnels ?	Moi même	Un service	Une personne											
2301			1	2	3											
2399	Q302=1	VIRAL														
Q24		VIRAL														
Q25		En général, vous partez en voyage	Pendant les	Hors vacances	Cela dépend.											
2501			1	2	3											
Q26	Q201<>(5,6,7)	Dans quel secteur travaillez-vous ? *(menu déroulant)*	Agriculture /	Architecture /	Automobile /	Energie / Chimi	Finance / Bang	Droit / Juridi	Conseil / Etude	Média / Com	Informatique /	Santé / Phar	Transport / Logis	Grande distri	Commerces et	Loisirs / Touri	Arts / Culture	Enseignement /	Administration /	Autre
2601			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
2699	Q201 IN (5,6,7)	VIRAL																		
Q27		Avez-vous déjà acheté des voyages sur les sites suivants :	Oui	Non																
2701		Accorhotel	1	0																
2702		Anyway.com	1	0																
2703		Bourse des voyages	1	0																
2704		Ebookers	1	0																
2705		Govoyages.com	1	0																
2706		Havas Voyages	1	0																
2707		Karavel.com	1	0																
2708		lastminute.com	1	0																
2709		Nouvelles Frontières	1	0																
2710		Opodo.fr	1	0																
2711		Partirdiscount.com	1	0																
2712		Promovacances.com	1	0																
2713		Travelprice.com	1	0																
2714		Vivacances.fr	1	0																
2715		voyages-sncf.com	1	0																

Q28		VIRAL

Q29		Si on vous offrait une nuit d'hôtel, dans quelle ville réserveriez-vous?
2901		menu déroulant, voir onglet correspondant

Q30		Disposez-vous d'une connexion Internet à votre domicile?	oui, bas débit	oui, haut débit	oui, mais ne connais pas le débit	je n'ai pas de connexion Internet à mon domicile
3001			1	2	3	4

Q31		Combien dépense votre foyer par an pour ses vacances ?	Moins de 762 € (5 000 F)	De 762 € (5 000F) à 1524 € (10 000 F)	De 1524 € (10 000 F) à 3048 € (20 000 F)	Plus de 3048 € (20 000 F)	Je ne souhaite pas répondre
3101			1	2	3	4	5

Q32		VIRAL

Q33		Y a-t-il des règles générales édictées par votre entreprise qui conditionnent :	Oui	Non
3301		Le choix de la compagnie	1	0
3302		La classe de voyage	1	0

Q34		A quelle fréquence visitez-vous les sites suivants :	Vous n'y êtes jamais allé	Moins d'une fois par mois	Une fois par mois ou plus
3401		Accorhotel	1	2	3
3402		Anyway.com	1	2	3
3403		Bourse des voyages	1	2	3
3404		Ebookers	1	2	3
3405		Govoyages.com	1	2	3
3406		Havas Voyages	1	2	3
3407		Karavel.com	1	2	3
3408		lastminute.com	1	2	3
3409		Nouvelles Frontières	1	2	3
3410		Opodo.fr	1	2	3
3411		Partirdiscount.com	1	2	3
3412		Promovacances.com	1	2	3
3413		Travelprice.com	1	2	3
3414		Vivacances.fr	1	2	3
3415		voyages-sncf.com	1	2	3

Q35		Avez-vous déjà choisi un billet d'avion uniquement parce qu'il était en promotion ?	Oui	Non	Ne sais pas
3501			1	0	2

Q36		VIRAL

Q37		Vous vivez dans...	Un appart	Une maiso	Une maison avec jardin
3701			1	2	3

Q38	Q101<>5	Vous êtes...	Propriétai	Locataire
3801			1	2
3899	Q101=5	VIRAL		

Q39		Quel est le niveau d'étude le plus haut que vous ayez atteint?	Collège	Lycée	Bac +1/2	Bac + 3/4	Bac + 5 et plus
3901			1	2	3	4	5

Q40		VIRAL

Q41		Sur Internet, avez-vous déjà acheté? (plusieurs choix possibles)
4101		Du matériel ou des logiciels informatiques
4102		Du matériel hifi/photo
4103		Des billets de train, des location de voitures
4104		Des voyages : séjours, thalassos, croisièrs,...
4105		Des produits culturels (Livres, CD audio ou DVD)
4106		Des billets pour des sorties, des spectacles
4107		Des vêtements
4108		Des produits de beauté
4109		Des fleurs ou des plantes
4110		Des produits alimentaires (courses en ligne)
4111		Du vin
4112		Des produits ou des services financiers
4113		D'autres produits
4114		Je n'ai jamais acheté sur Internet

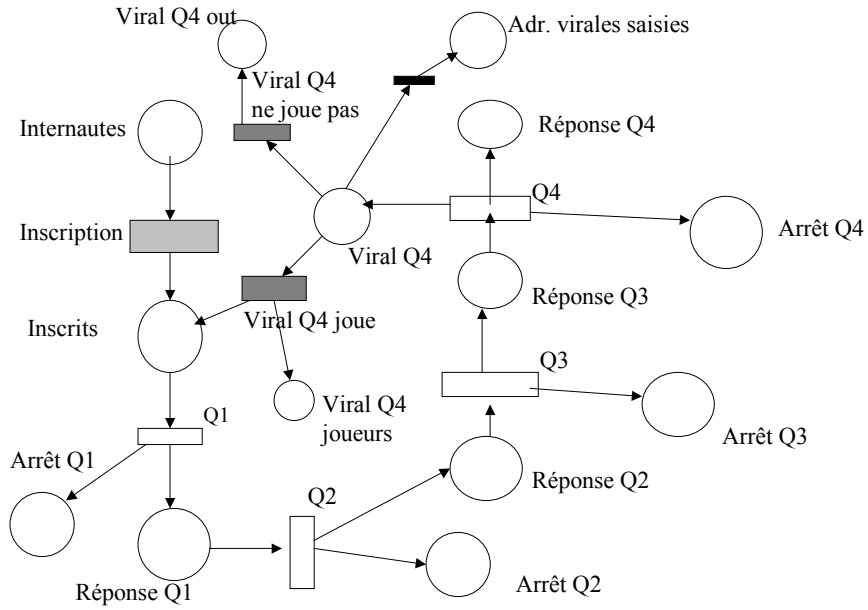
L'ensemble des résultats de l'opération « Le Ciel est à tout le monde » montre que ce jeu a été particulièrement performant tout en ayant répondu aux objectifs de la société Anyway.
Plusieurs questions cependant se posent : aurait-on récupéré plus d'adresses e-mail par un design de questionnaire différent ? Comment maximiser le nombre d'adresses e-mail récupérées par ce jeu en optimisant le nombre et l'ordre des questions virales et des questions de qualification ? Comment mesurer l'efficacité globale du jeu en terme de marketing viral ? Existe-t-il différents types d'internautes : les bons « viralisateurs » et les moins bons ? Nous définissons ici les « viralisateurs » comme des individus qui d'une part fournissent de nombreuses adresses virales de qualité (la qualité portant sur l'adresse e-mail (correcte ou non) et qui d'autre part donnent de bons « filleuls », les filleuls correspondant aux adresses virales fournies et transformées en joueurs et qui sont de plus des prospects intéressants pour l'entreprise sur la base de leurs comportements de voyage (fournis dans les questions de qualification)). Les réponses à toutes ces questions marketing exigent d'avoir recours à des techniques de modélisation qui vont nous permettre de répondre à ces problématiques. Nous proposons donc dans ce chapitre d'utiliser les réseaux de Petri pour représenter l'ensemble du système de ce jeu promotionnel viral.

5.4 MODELISATION DU JEU PROMOTIONNEL « LE CIEL EST A TOUT LE MONDE » PAR LES RESEAUX DE PETRI :

L'opération « Le ciel est à tout le monde » peut être décomposée en un certain nombre d'actions séquentielles effectuées par l'internaute, selon le mécanisme décrit précédemment (figure 5.1). Ces actions correspondent à différents évènements sur le système et peuvent donc être représentées par un réseau de Petri. Afin de comprendre clairement la manière dont nous allons modéliser ce jeu, nous allons commencer par un exemple simple, la description du jeu ne comportant que les 4 premières questions du questionnaire.

Le jeu peut être représenté par le réseau de Petri suivant :

Figue 5.2: Réseau de Pétri de l'opération « Le ciel est à tout le monde » : exemple des 4 premières questions.



La place initiale du système est la place *Internautes* qui correspond aux internautes qui viennent sur le site du jeu promotionnel. La transition *inscription* s'effectue lorsqu'un joueur de la place *joueurs* s'inscrit au jeu. Le joueur inscrit est alors placé dans la place *Inscrits.* Il est ensuite soumis à la question Q1, représentée par la transition *Q1* sur la figure 5.2. Si le joueur répond à la question Q1, il passe dans la place *réponseQ1.* En revanche, s'il s'arrête de jouer sans répondre à la question Q1, il passe dans la place *ArrêtQ1*. Dans le cas où le joueur a répondu à la question Q1, le jeu se poursuit : il est ensuite soumis à la question Q2 et ainsi de suite. Lorsque le joueur arrive à une question virale, comme la Q4 par exemple, il peut soit répondre (place *RéponseQ4*) ou ne pas répondre (place *ArrêtQ4*), mais on va également pour cette question virale mesurer le nombre d'adresses virales fournies. En effet, chaque joueur répondant à la question Q4 passe par la transition Q4 qui crée un jeton dans la place *ViralQ4*, ce qui indique que ce joueur a fourni une adresse e-mail. L'adresse virale ainsi saisie reçoit un e-mail l'invitant à participer au jeu (ceci n'est pas représenté sur le réseau de Petri car c'est une action immédiate et systématique n'influençant pas les paramètres du modèle. De plus, le rajout de places et transitions pour décrire cet envoi d'e-mail compliquerait le graphique et diminuerait sa compréhension). Lorsqu'il reçoit l'e-mail d'invitation au jeu, l'internaute dont le joueur a fourni l'adresse virale a deux possibilités: soit il joue,

soit il ne joue pas. S'il joue, il passe par la transition *ViralQ4joue* puis dans la place *Inscrits* et rentre dans le mécanisme du jeu. On comptabilise le nombre d'adresses virales qui sont devenues joueurs (indicateur de transformation de viral en joueur) dans la place *ViralQ4joueurs*. S'il ne joue pas, il passe par la transition *ViralQ4nejouepas* puis dans la place *ViralQ4out* qui fournit donc le nombre d'adresses virales fournies n'ayant pas joué.

Comme on peut le constater dans le questionnaire du tableau 5.1, il existe trois catégories de questions: les questions de qualification « pures » (demande de critères socio-démographiques ou comportement de voyage); les questions virales « pures » (demande d'e-mail d'un ami); les questions « mixtes » qui sont soit de qualification soit virales en fonction d'une question filtre préalable. Un exemple de question mixte est la Q13: la question de qualification posée est « Pour quelle(s) raison(s) n'achetez-vous pas vos voyages personnels sur Internet ? ». Les participants ayant répondu n'avoir jamais acheté sur Internet des voyages personnels (question Q9 filtre) reçoivent en Q13 la question de qualification (notée Q13Q) tandis que ceux ayant répondu avoir acheté des voyages personnels sur Internet en Q9 reçoivent en Q13 une question virale (donner l'adresse e-mail d'un ami) notée Q13V.

Le questionnaire du jeu comporte 41 questions, dont 25 questions de qualification « pure », 9 questions virales « pures » (Q4, Q8, Q16, Q20, Q24, Q28, Q32, Q36, Q40) et 7 questions « mixtes » (Q13, Q15, Q17, Q18, Q23, Q26, Q38).

Les indicateurs de performance du jeu tirés du réseau de Petri de la figure 5.2 sont par conséquent :

- **Le nombre d'adresses virales fournies par le jeu (nombre de jetons de la place *Adr.Virales saisies*).**

  Précisons que la place *Adr.Virales saisies* est une place de comptage permettant de calculer le nombre d'adresses virales saisies. Elle est reliée à la transition *ViralQ4* par une transition immédiate (rectangle noir) : à chaque fois qu'une adresse virale sera fournie en Q4, un jeton sera d'une part crée dans la place *Adr.virales saisies* (pour compter le nombre d'adresses virales) et d'autre part, de manière parallèle, un jeton ira soit vers la transition *ViralQ4joue*, soit vers *ViralQ4nejouepas* en fonction des probabilités de ces évènements.
- **Le nombre d'adresses virales ayant joué (nombre de jetons de la place *ViralQ4joueurs*)**

Là encore, la place *ViralQ4joueurs* est dans le réseau une place de comptage du nombre d'adresses virales ayant joué. Lorsqu'une adresse virale joue (transition *ViralQ4joue*), un jeton va automatiquement dans la place *ViralQ4joueurs* et un autre jeton va parallèlement dans la place *Inscrits*.

- **Le nombre de non réponse par question (nombre de jetons des places *ArrêtQ1, ArrêtQ2, ArrêtQ3, ArrêtQ4*).**

Les réseaux de Petri stochastiques reposant sur des distributions de probabilités définissant chacune des transitions du système, nous allons donc modéliser les différentes lois des transitions de la figure 5.1 et introduire également la notion de temps dans le modèle. Nous devons donc dans un premier temps déterminer un modèle pour chaque transition du réseau de Petri, à savoir :

- **Loi de la transition *Inscription* : l'objectif est de modéliser le nombre de joueurs qui s'inscrivent au jeu pour chacun des 30 jours du jeu.**
- **Loi des transitions *Q1, Q2, Q3, Q4* : l'objectif est de modéliser la probabilité et les délais de réponse (ou de non réponse) pour chaque question.**
- **Loi de la transition *ViralQ4joue* : si on généralise à l'ensemble des questions virales du questionnaire, l'objectif est ici de modéliser la probabilité pour chaque question virale que l'adresse virale joue, ainsi que la durée qui s'écoule entre le moment auquel l'adresse virale est saisie par un joueur et le moment auquel l'adresse virale saisie s'inscrit au jeu.**

Les paramètres de ces différents modèles ainsi que les indicateurs de performances ci-dessus décrits seront ensuite programmés dans le réseau de Petri afin d'avoir un modèle représentant le système global et permettant d'effectuer des simulations.

5.4.1 Modèle du nombre d'inscriptions de joueurs par jour :

L'objectif de cette partie est de déterminer la loi de la transition *inscription* dans le réseau de Petri du jeu Anyway de la figure 5.1. Ceci revient à modéliser le nombre d'inscriptions de joueurs pour chacun des 30 jours de l'opération.

Les données observées sont les suivantes :

Tableau 5.3 : Nbre d'inscriptions de joueurs par jour

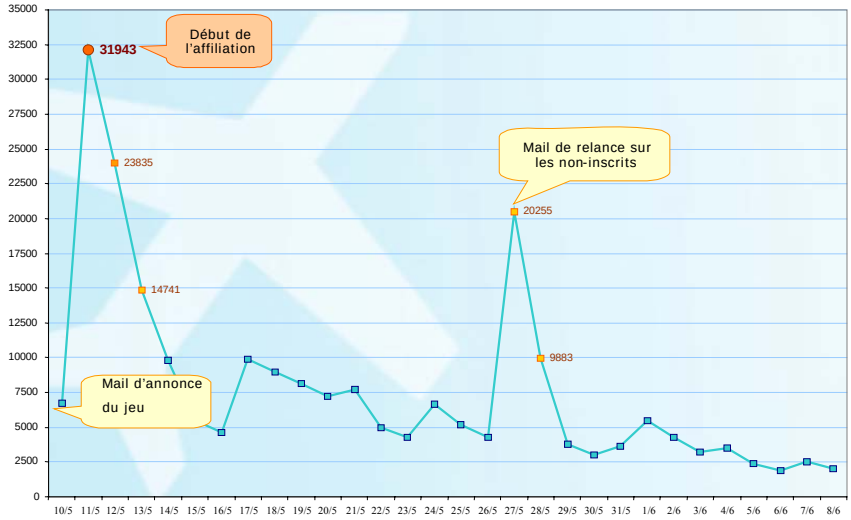
JOUR * JOUEUR Crosstabulation

Count

		JOUEUR	
		Selected	Total
JOUR	10-MAY-2004	6700	6700
	11-MAY-2004	31943	31943
	12-MAY-2004	23835	23835
	13-MAY-2004	14741	14741
	14-MAY-2004	9682	9682
	15-MAY-2004	5370	5370
	16-MAY-2004	4510	4510
	17-MAY-2004	9791	9791
	18-MAY-2004	8891	8891
	19-MAY-2004	8022	8022
	20-MAY-2004	7158	7158
	21-MAY-2004	7648	7648
	22-MAY-2004	4942	4942
	23-MAY-2004	4166	4166
	24-MAY-2004	6587	6587
	25-MAY-2004	5087	5087
	26-MAY-2004	4153	4153
	27-MAY-2004	20255	20255
	28-MAY-2004	9883	9883
	29-MAY-2004	3693	3693
	30-MAY-2004	2902	2902
	31-MAY-2004	3531	3531
	01-JUN-2004	5345	5345
	02-JUN-2004	4221	4221
	03-JUN-2004	3096	3096
	04-JUN-2004	3426	3426
	05-JUN-2004	2262	2262
	06-JUN-2004	1742	1742
	07-JUN-2004	2392	2392
	08-JUN-2004	1905	1905
Total		227879	227879

Ces données observées jour par jour sont représentées sur la figure 5.3. On constate que le nombre d'inscriptions varie selon les jours de l'opération et que les forts pics d'inscription (premiers jours du jeu et autour du 18ème jour) sont a priori liés à des actions marketing menées par la société Millemercis pour recruter des joueurs (affiliation, mail d'annonce, mail de relance, autopromotion sur le site Anyway).

Figure 5.3: Nombre d'inscrits au jour le jour



Les inscrits au jeu proviennent de cinq sources : l'affiliation (avec d'autres sites Web), le marketing viral (adresses données par des déjà participants au jeu), le mail d'annonce du jeu sur la base de clients Anyway, le mail de relance sur les clients de la base Anyway qui n'avaient pas joué après le mail d'annonce, l'autopromotion sur le site anyway.

En plus de ces sources de joueurs, un certain nombre d'internautes viennent jouer directement sans être soumis à un stimulus quelconque.

Les chiffres de recrutement par jour et par source sont les suivants :

Tableau 5.4.1: Source de provenance des joueurs

Jour	inscription directe	affiliation	viral	mail annonce	autopromo site	mail relance	Total	%
10 5	127	142	521	5910			6700	2,94
11 5	753	9688	7247	14255			31943	14,02
12 5	650	13943	6649	2593			23835	10,46
13 5	402	8050	5092	1053	144		14741	6,47
14 5	160	5494	3353	486	189		9682	4,25
15 5	81	3065	1827	288	109		5370	2,36
16 5	74	2548	1511	253	124		4510	1,98
17 5	157	6062	2973	344	255		9791	4,30
18 5	158	5435	2886	179	233		8891	3,90
19 5	146	5097	2480	128	171		8022	3,52
20 5	114	4690	2183	75	96		7158	3,14
21 5	131	4594	2695	99	129		7648	3,36
22 5	89	2873	1856	52	72		4942	2,17
23 5	70	2270	1675	66	85		4166	1,83
24 5	117	3402	2901	79	88		6587	2,89
25 5	101	2461	2295	47	180		5084	2,23
26 5	79	2194	1681	36	163		4153	1,82
27 5	351	4884	3650	40	233	11097	20255	8,89
28 5	157	2841	3255	15	171	3444	9883	4,34
29 5	76	1365	1307	12	75	858	3693	1,62
30 5	67	1167	1070	13	74	511	2902	1,27
31 5	86	1508	1214	12	107	604	3531	1,55
1 6	109	2123	2117	12	121	863	5345	2,35
2 6	108	2010	1544	10	148	401	4221	1,85
3 6	81	1591	1144	14	36	230	3096	1,36
4 6	73	2079	1025	7	97	145	3426	1,50
5 6	52	1378	675	7	64	86	2262	0,99
6 6	42	999	572	9	60	60	1742	0,76
7 6	198	1409	615	11	76	83	2392	1,05
8 6	64	1064	618	5	95	59	1905	0,84
TOTAL	4873	106426	68631	26110	3395	18441	227876	100,00
%	2,14	46,70	30,12	11,46	1,49	8,09	100,00	

On constate sur le tableau 5.4.1 que sur l'ensemble des 30 jours du jeu, la principale source de recrutement pour le jeu est l'affiliation (47% des joueurs viennent de sites affiliés), suivie du viral (30% des joueurs) puis du mail d'annonce sur les clients Anyway (11%) et du mail de relance (8%). Les autres sources (inscription directe et autopromotion sur le site) sont des sources de recrutement très faibles. En revanche, comme le montre le tableau 5.4.2, le poids des sources de recrutement diffère pour chaque jour du jeu, comme par exemple pour le 18ème jour du jeu pour lequel 54,8% du recrutement provient du mail de relance et 24,1% de l'affiliation.

Tableau 5.4.2: Poids des sources de provenance des joueurs par jour

		Jeu direct (ss source identifiée)	% par jour	affiliation	% par jour	viral	% par jour	mail annonce	% par jour	autopromo site	% par jour	mail relance	% par jour	TOTAL	
1	10 5	127	1,90	142	2,12	521	7,78	5910	88,21					6700	100%
2	11 5	753	2,36	9688	30,33	7247	22,69	14255	44,63					31943	100%
3	12 5	650	2,73	13943	58,50	6649	27,90	2593	10,88					23835	100%
4	13 5	402	2,73	8050	54,61	5092	34,54	1053	7,14	144	0,98			14741	100%
5	14 5	160	1,65	5494	56,74	3353	34,63	486	5,02	189	1,95			9682	100%
6	15 5	81	1,51	3065	57,08	1827	34,02	288	5,36	109	2,03			5370	100%
7	16 5	74	1,64	2548	56,50	1511	33,50	253	5,61	124	2,75			4510	100%
8	17 5	157	1,60	6062	61,91	2973	30,36	344	3,51	255	2,60			9791	100%
9	18 5	158	1,78	5435	61,13	2886	32,46	179	2,01	233	2,62			8891	100%
10	19 5	146	1,82	5097	63,54	2480	30,91	128	1,60	171	2,13			8022	100%
11	20 5	114	1,59	4690	65,52	2183	30,50	75	1,05	96	1,34			7158	100%
12	21 5	131	1,71	4594	60,07	2695	35,24	99	1,29	129	1,69			7648	100%
13	22 5	89	1,80	2873	58,13	1856	37,56	52	1,05	72	1,46			4942	100%
14	23 5	70	1,68	2270	54,49	1675	40,21	66	1,58	85	2,04			4166	100%
15	24 5	117	1,78	3402	51,65	2901	44,04	79	1,20	88	1,34			6587	100%
16	25 5	101	1,99	2461	48,38	2295	45,11	47	0,92	180	3,54			5087	100%
17	26 5	79	1,90	2194	52,83	1681	40,48	36	0,87	163	3,92			4153	100%
18	27 5	351	1,73	4884	24,11	3650	18,02	40	0,20	233	1,15	11097	54,79	20255	100%
19	28 5	157	1,59	2841	28,75	3255	32,94	15	0,15	171	1,73	3444	34,85	9883	100%
20	29 5	76	2,06	1365	36,96	1307	35,39	12	0,32	75	2,03	858	23,23	3693	100%
21	30 5	67	2,31	1167	40,21	1070	36,87	13	0,45	74	2,55	511	17,61	2902	100%
22	31 5	86	2,44	1508	42,71	1214	34,38	12	0,34	107	3,03	604	17,11	3531	100%
23	1 6	109	2,04	2123	39,72	2117	39,61	12	0,22	121	2,26	863	16,15	5345	100%
24	2 6	108	2,56	2010	47,62	1544	36,58	10	0,24	148	3,51	401	9,50	4221	100%
25	3 6	81	2,62	1591	51,39	1144	36,95	14	0,45	36	1,16	230	7,43	3096	100%
26	4 6	73	2,13	2079	60,68	1025	29,92	7	0,20	97	2,83	145	4,23	3426	100%
27	5 6	52	2,30	1378	60,92	675	29,84	7	0,31	64	2,83	86	3,80	2262	100%
28	6 6	42	2,41	999	57,35	572	32,84	9	0,52	60	3,44	60	3,44	1742	100%
29	7 6	198	8,28	1409	58,90	615	25,71	11	0,46	76	3,18	83	3,47	2392	100%
30	8 6	64	3,36	1064	55,85	618	32,44	5	0,00	95	4,99	59	3,10	1905	100%

Comme le montre le tableau 5.4.1, les pics d'inscriptions se situent au 2ème jour du jeu (14% des joueurs s'inscrivent sur cette journée), au 3ème jour (10%) et au 18ème jour du jeu (9% des joueurs).

Etant donné que nous souhaitons modéliser le nombre d'inscriptions de joueurs par jour pour paramétrer dans le réseau de Petri la loi de la transition *inscription*, nous allons modéliser, pour chaque source de recrutement et par jour, le nombre d'inscriptions. Ceci nous permettra ensuite d'effectuer des simulations de stratégies marketing différentes grâce au réseau de Petri, et de répondre au type de question suivante :

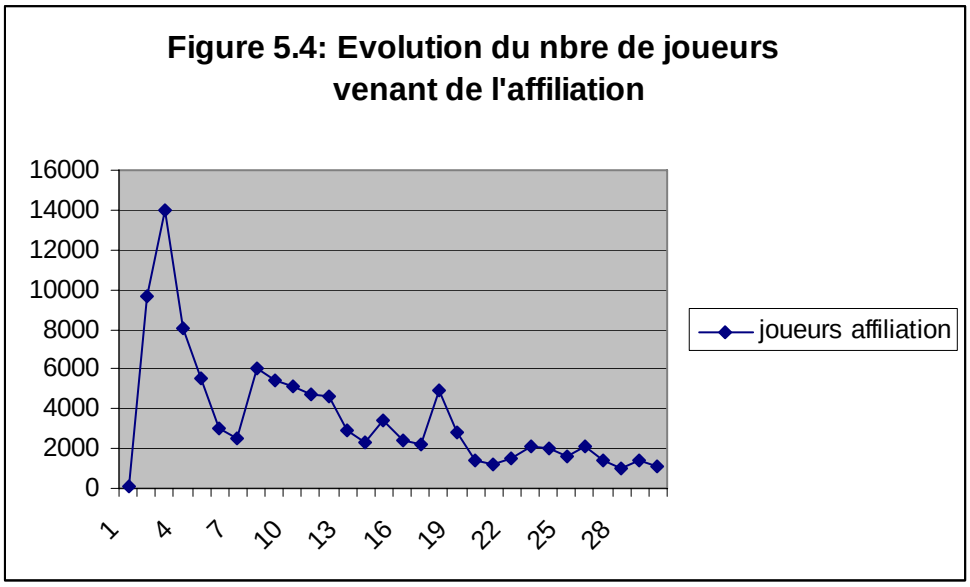
- Comment évolue le nombre de joueurs selon qu'on fait plus d'affiliation, de mails d'annonce ou de relance, ou qu'on favorise l'effet viral du jeu ?
- Quelle est le meilleur mix de sources de recrutement (entre affiliation, viral, mailing, etc…) ?

Nous ne modéliserons pas le nombre d'inscriptions liées au marketing viral puisque cette variable est endogène au système et sera modélisée directement dans le réseau de Petri.

Par conséquent, nous allons ci-dessous analyser chaque source de recrutement exogène (à savoir toutes sauf le viral), en commençant par les plus importantes.

5.4.1.1 Modélisation du nombre d'inscriptions venant de l'affiliation

Nous cherchons dans cette partie à modéliser le nombre de joueurs recrutés par l'affiliation. L'affiliation est une stratégie couramment utilisée par les entreprises qui se développent sur Internet. Elle consiste à avoir sur des sites partenaires un lien vers son propre site de manière à générer du trafic, les sites partenaires tant généralement des sites à forte audience. Dans la majorité des cas, le site rémunère le site partenaire par un pourcentage du CA généré par les internautes venant du site partenaire (entre 5 et 8% environ).
Comme le montre le tableau 5.4 (source de provenance des joueurs), le nombre d'inscrits provenant de l'affiliation varie d'un jour à l'autre avec des pics aux 2ème, 3ème et 4ème jours du jeu. La figure 5.4 ci-dessous représente le nombre de joueurs par jour recrutés par l'affiliation (en raison de l'échelle allant de 0 à 16000, on pourrait supposer que la première donnée est à zéro mais le nombre de joueurs recrutés par affiliation le premier jour est en réalité 142).



Nous avons testé tout d'abord pour modéliser ce phénomène une loi exponentielle, ce qui est cohérent avec de nombreuses recherches sur les files d'attentes où la loi d'arrivée des individus ou d'un processus est souvent une loi exponentielle (Trivedi, 2002). Ceci est également cohérent avec la propriété de la distribution exponentielle qui est d'être sans mémoire (les inscriptions des joueurs sont indépendantes).

Modèle exponentiel :

Rappelons qu'une distribution exponentielle est de la forme suivante :

$$f(X) = \lambda e^{-\lambda X}$$

λ paramètre du modèle.

La fonction de répartition de la loi exponentielle est :

$$F(x) = P(T \leq x) = 1 - e^{-\lambda x}$$

En utilisant le solveur d'Excel (Hardie et Fader, 2002) et en appliquant la méthode du maximum de vraisemblance, on estime le paramètre lambda de la loi exponentielle.

Rappel de la méthode du maximum de vraisemblance :

La fonction de vraisemblance est définie comme la probabilité d'observer tous les points de données. Cette probabilité est calculée à partir du modèle, comme fonction des paramètres du modèle :

$$L(\text{paramètres}) = p(\text{données/paramètres})$$

Pour un ensemble de paramètres donné, L (paramètres) donne la probabilité d'obtenir les données réelles. Pour un jeu de données, les paramètres estimés du modèle sont ceux qui maximisent cette vraisemblance.

L'application de cette méthode pour modéliser le nombre d'inscriptions de joueurs par jour issu de l'affiliation donne les résultats suivants (les formules utilisées pour calculer les paramètres du modèle exponentiel sont fournies dans le tableau 5.5 ; les résultats sont présentés dans le tableau 5.6).

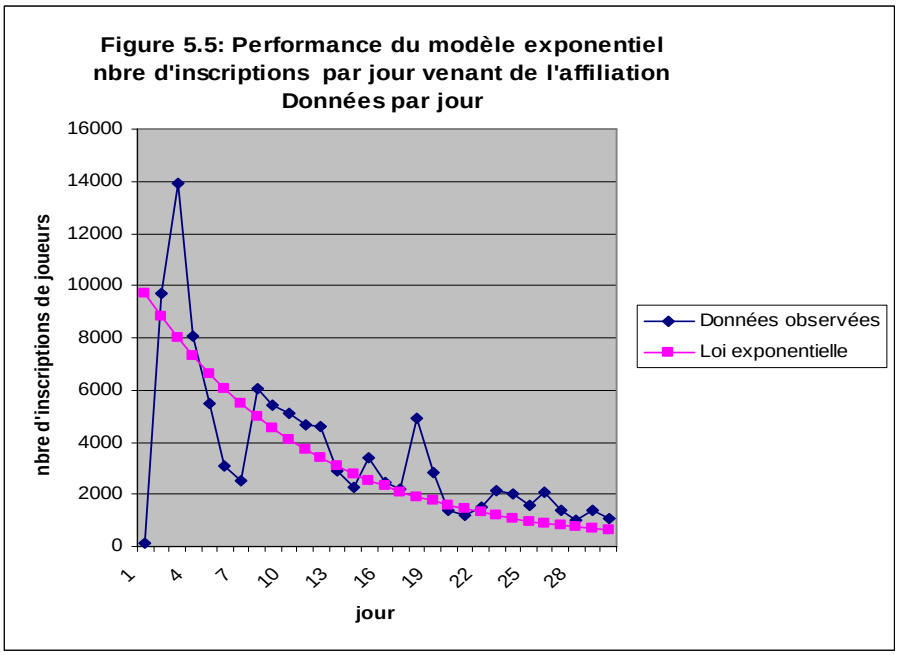
Tableau 5.5 : Paramètres de la loi exponentielle calculés sous Excel (Formules)

Nbre de joueurs 106426				lambda LL=	0,0958064802985758 =SOMME(H5:H34)	
	Jour	Cum incscriptions	Incremental	P(T<=t)	LL	E(T(t))
1	38117	=E5	142	=1-EXP(-H$1*A5)	=E5*LN(G5)	=B$2*G5
=A5+1	38118	=E6+D5	9688	=1-EXP(-H$1*A6)	=E6*LN(G6-G5)	=B$2*G6
=A6+1	38119	=E7+D6	13943	=1-EXP(-H$1*A7)	=E7*LN(G7-G6)	=B$2*G7
=A7+1	38120	=E8+D7	8050	=1-EXP(-H$1*A8)	=E8*LN(G8-G7)	=B$2*G8
=A8+1	38121	=E9+D8	5494	=1-EXP(-H$1*A9)	=E9*LN(G9-G8)	=B$2*G9
=A9+1	38122	=E10+D9	3065	=1-EXP(-H$1*A10)	=E10*LN(G10-G9)	=B$2*G10
=A10+1	38123	=E11+D10	2548	=1-EXP(-H$1*A11)	=E11*LN(G11-G10)	=B$2*G11
=A11+1	38124	=E12+D11	6062	=1-EXP(-H$1*A12)	=E12*LN(G12-G11)	=B$2*G12
=A12+1	38125	=E13+D12	5435	=1-EXP(-H$1*A13)	=E13*LN(G13-G12)	=B$2*G13
=A13+1	38126	=E14+D13	5097	=1-EXP(-H$1*A14)	=E14*LN(G14-G13)	=B$2*G14
=A14+1	38127	=E15+D14	4690	=1-EXP(-H$1*A15)	=E15*LN(G15-G14)	=B$2*G15
=A15+1	38128	=E16+D15	4594	=1-EXP(-H$1*A16)	=E16*LN(G16-G15)	=B$2*G16
=A16+1	38129	=E17+D16	2873	=1-EXP(-H$1*A17)	=E17*LN(G17-G16)	=B$2*G17
=A17+1	38130	=E18+D17	2270	=1-EXP(-H$1*A18)	=E18*LN(G18-G17)	=B$2*G18
=A18+1	38131	=E19+D18	3402	=1-EXP(-H$1*A19)	=E19*LN(G19-G18)	=B$2*G19
=A19+1	38132	=E20+D19	2461	=1-EXP(-H$1*A20)	=E20*LN(G20-G19)	=B$2*G20
=A20+1	38133	=E21+D20	2194	=1-EXP(-H$1*A21)	=E21*LN(G21-G20)	=B$2*G21
=A21+1	38134	=E22+D21	4884	=1-EXP(-H$1*A22)	=E22*LN(G22-G21)	=B$2*G22
=A22+1	38135	=E23+D22	2841	=1-EXP(-H$1*A23)	=E23*LN(G23-G22)	=B$2*G23
=A23+1	38136	=E24+D23	1365	=1-EXP(-H$1*A24)	=E24*LN(G24-G23)	=B$2*G24
=A24+1	38137	=E25+D24	1167	=1-EXP(-H$1*A25)	=E25*LN(G25-G24)	=B$2*G25
=A25+1	38138	=E26+D25	1508	=1-EXP(-H$1*A26)	=E26*LN(G26-G25)	=B$2*G26
=A26+1	38139	=E27+D26	2123	=1-EXP(-H$1*A27)	=E27*LN(G27-G26)	=B$2*G27
=A27+1	38140	=E28+D27	2010	=1-EXP(-H$1*A28)	=E28*LN(G28-G27)	=B$2*G28
=A28+1	38141	=E29+D28	1591	=1-EXP(-H$1*A29)	=E29*LN(G29-G28)	=B$2*G29
=A29+1	38142	=E30+D29	2079	=1-EXP(-H$1*A30)	=E30*LN(G30-G29)	=B$2*G30
=A30+1	38143	=E31+D30	1378	=1-EXP(-H$1*A31)	=E31*LN(G31-G30)	=B$2*G31
=A31+1	38144	=E32+D31	999	=1-EXP(-H$1*A32)	=E32*LN(G32-G31)	=B$2*G32
=A32+1	38145	=E33+D32	1409	=1-EXP(-H$1*A33)	=E33*LN(G33-G32)	=B$2*G33
=A33+1	38146	=E34+D33	1064	=1-EXP(-H$1*A34)	=E34*LN(G34-G33)	=B$2*G34

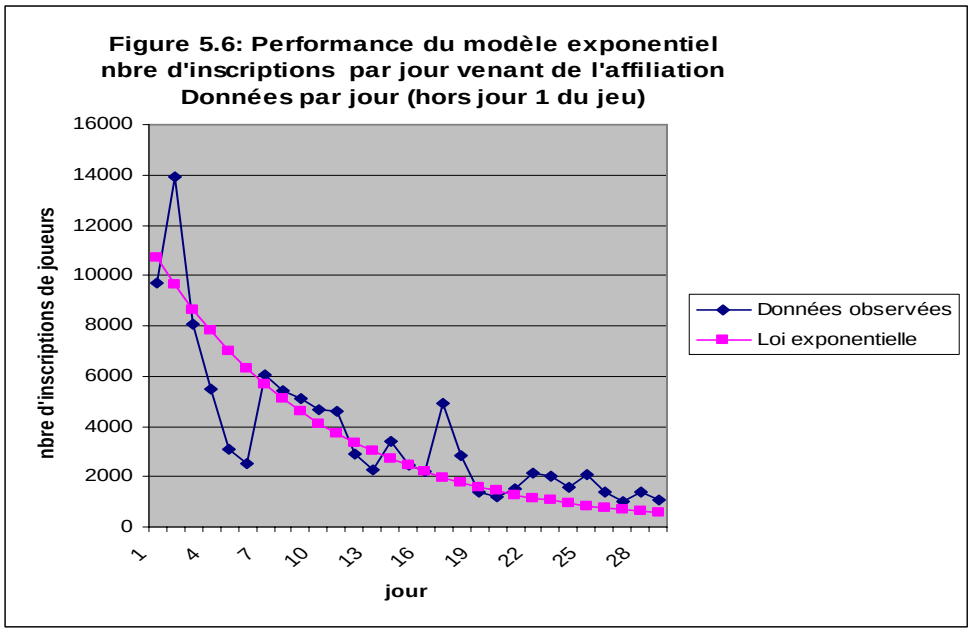
Tableau 5.6 : Paramètres de la loi exponentielle calculés sous Excel (Résultats)

Nbre de joueurs 106426				lambda LL=	0,0958 -356080,8896	
	Jour	Cum incscriptions	Incremental	P(T<=t)	LL	E(T(t))
1	10/5/04	142	142	0,0914	-339,7983	9723,0965
2	11/5/04	9830	9688	0,1744	-24111,0318	18557,8894
3	12/5/04	23773	13943	0,2498	-36036,5024	26585,5341
4	13/5/04	31823	8050	0,3183	-21576,9399	33879,7720
5	14/5/04	37317	5494	0,3806	-15252,2873	40507,6071
6	15/5/04	40382	3065	0,4372	-8802,6131	46529,9221
7	16/5/04	42930	2548	0,4886	-7561,9153	52002,0374
8	17/5/04	48992	6062	0,5353	-18571,4894	56974,2194
9	18/5/04	54427	5435	0,5778	-17171,3260	61492,1420
10	19/5/04	59524	5097	0,6164	-16591,7753	65597,3066
11	20/5/04	64214	4690	0,6514	-15716,2396	69327,4226
12	21/5/04	68808	4594	0,6833	-15834,6775	72716,7546
13	22/5/04	71681	2873	0,7122	-10177,9574	75796,4367
14	23/5/04	73951	2270	0,7385	-8259,2361	78594,7585
15	24/5/04	77353	3402	0,7624	-12703,8725	81137,4252
16	25/5/04	79814	2461	0,7841	-9425,7358	83447,7935
17	26/5/04	82008	2194	0,8038	-8613,3137	85547,0862
18	27/5/04	86892	4884	0,8217	-19641,7676	87454,5871
19	28/5/04	89733	2841	0,8380	-11697,7107	89187,8184
20	29/5/04	91098	1365	0,8528	-5751,1120	90762,7015
21	30/5/04	92265	1167	0,8663	-5028,6910	92193,7030
22	31/5/04	93773	1508	0,8785	-6642,5618	93493,9679
23	1/6/04	95896	2123	0,8896	-9554,9613	94675,4404
24	2/6/04	97906	2010	0,8997	-9238,9546	95748,9735
25	3/6/04	99497	1591	0,9088	-7465,4514	96724,4283
26	4/6/04	101576	2079	0,9172	-9954,4761	97610,7655
27	5/6/04	102954	1378	0,9247	-6730,0338	98416,1267
28	6/6/04	103953	999	0,9316	-4974,7410	99147,9100
29	7/6/04	105362	1409	0,9379	-7151,4178	99812,8375
30	8/6/04	106426	1064	0,9435	-5502,2991	100417,0171

La performance du modèle exponentiel par rapport aux données observées est présentée sous forme graphique par la figure 5.5. On constate globalement une approximation des données par le modèle mais avec de nombreux écarts. Dans les données observées du nombre d'inscriptions de joueurs par jour (tableau 5.1), on observe quelques pics particuliers, notamment en début de jeu (les 11, 12 et 13 mai) qui sont mal représentés dans un modèle exponentiel.



On constate notamment un écart sur le premier jour, pour lequel le nombre de joueurs observés est très faible (142) par rapport aux observations sur les autres jours du jeu alors que le modèle prévoit 9723 joueurs sur ce jour. Ceci provient d'une politique d'affiliation qui s'est mise en place à différents moments de la journée du 10 Mai et qui a eu son impact essentiellement sur le 2ème jour du jeu. Etant donné que nous ne possédons pas le détail horaire des mises en place de l'affiliation, nous avons fait l'hypothèse que l'affiliation s'est mise en place en fin du premier jour. Nous avons donc appliqué le modèle exponentiel à partir du 2ème jour du jeu, dont les résultats sont présentés dans la figure 5.6.



Le modèle présenté ici de meilleurs résultats, malgré encore quelques pics de différences entre les données observées et estimées (entre le 4ème et le 7ème jour par exemple).

Afin d'améliorer le modèle, nous avons introduit une hétérogénéité entre les individus et testé un modèle exponentiel gamma dans lequel le paramètre lambda du modèle exponentiel varie selon les individus. En effet, on peut considérer que le nombre d'inscriptions journalier dépend du type de joueur qui s'inscrit, les joueurs s'inscrivant au début du jeu étant probablement différents de ceux qui s'inscrivent en fin de jeu.

Modèle Exponentiel-Gamma :

Le modèle Exponentiel Gamma est défini de la manière suivante :

- la probabilité de non réponse par question pour un joueur suit une loi exponentielle de la forme suivante :

$$f(X) = \lambda e^{-\lambda X}$$

λ paramètre du modèle.

(la fonction de répartition de la loi exponentielle est $F(x) = P(T \leq x) = 1 - e^{-\lambda x}$)

- il y a hétérogénéité entre les joueurs, ce qui se traduit par un paramètre lambda qui est distribué sur l'ensemble des joueurs selon une distribution gamma :

$$g(\lambda)=\frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda^r}}{\Gamma(r)}$$

où r est le paramètre de forme du modèle et α le paramètre d'échelle

- La distribution cumulée du modèle exponentiel-gamma est :

$$P(T\le t)=1-\left(\frac{\alpha}{\alpha+t}\right)^r$$

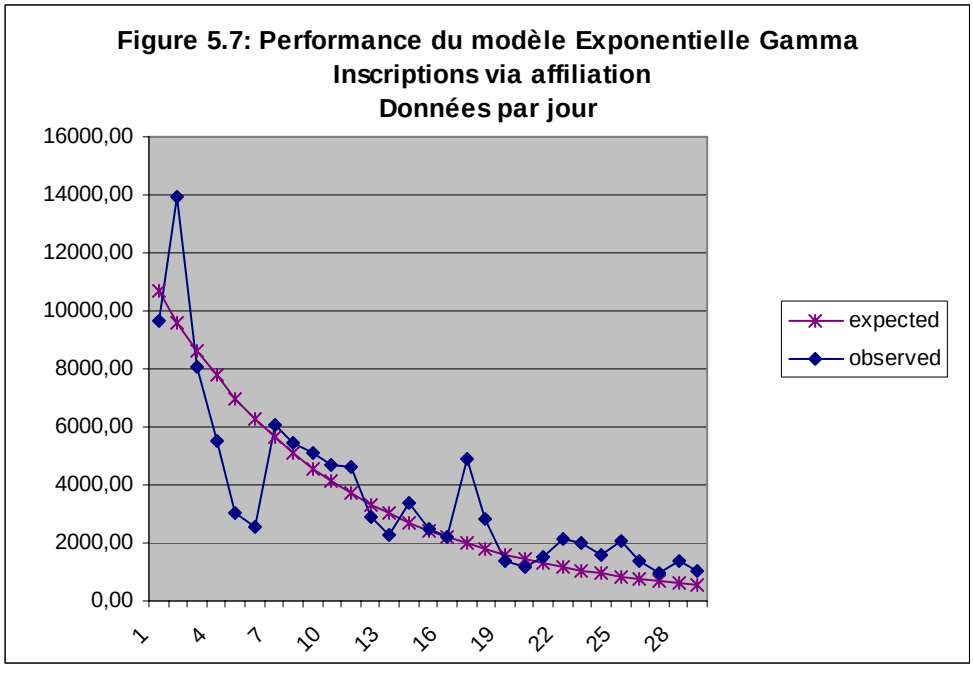
Les tableaux 5.7 et 5.8 donnent les formules et les résultats obtenus sous Excel pour le calcul des paramètres du modèle Exponentiel Gamma (par la méthode de maximisation de la vraisemblance). Comme pour le modèle exponentiel, nous n'avons modélisé l'affiliation qu'à partir du 2ème jour du jeu uniquement. On constate que la vraisemblance obtenue (-345050) n'est pas meilleure que le modèle exponentiel La performance du modèle est fournie par la figure 5.7 qui compare les données observées et les données estimées. On en conclut que le modèle Exponentiel Gamma n'apporte pas plus de précision pour représenter le nombre de joueurs venant de l'affiliation pour chaque jour du jeu.

Tableau 5.7 : Paramètres du modèle Exponentiel Gamma (formules)

Nbre jou	106426				r	61218510,1186534			
					alpha	578521940,357428			
					LL=	=SOMME(G7:G35)			
Questior	incr_Trl	Cumul inscrip	P(T<=t)		P(inscription)	LL	E(T(t))	PAR JO Observe	Expected
1	9688	=C7	=1-(G$2/(G$2+A7))^G$1		=E7	=C7*LN(F7)	=B$1*E7	=C7	=B$1*F7
2	13943	=C8+D7	=1-(G$2/(G$2+A8))^G$1		=E8-E7	=C8*LN(F8)	=B$1*E8	=C8	=B$1*F8
3	8050	=C9+D8	=1-(G$2/(G$2+A9))^G$1		=E9-E8	=C9*LN(F9)	=B$1*E9	=C9	=B$1*F9
4	5494	=C10+D9	=1-(G$2/(G$2+A10))^G$1		=E10-E9	=C10*LN(F10)	=B$1*E10	=C10	=B$1*F10
5	3065	=C11+D10	=1-(G$2/(G$2+A11))^G$1		=E11-E10	=C11*LN(F11)	=B$1*E11	=C11	=B$1*F11
6	2548	=C12+D11	=1-(G$2/(G$2+A12))^G$1		=E12-E11	=C12*LN(F12)	=B$1*E12	=C12	=B$1*F12
7	6062	=C13+D12	=1-(G$2/(G$2+A13))^G$1		=E13-E12	=C13*LN(F13)	=B$1*E13	=C13	=B$1*F13
8	5435	=C14+D13	=1-(G$2/(G$2+A14))^G$1		=E14-E13	=C14*LN(F14)	=B$1*E14	=C14	=B$1*F14
9	5097	=C15+D14	=1-(G$2/(G$2+A15))^G$1		=E15-E14	=C15*LN(F15)	=B$1*E15	=C15	=B$1*F15
10	4690	=C16+D15	=1-(G$2/(G$2+A16))^G$1		=E16-E15	=C16*LN(F16)	=B$1*E16	=C16	=B$1*F16
11	4594	=C17+D16	=1-(G$2/(G$2+A17))^G$1		=E17-E16	=C17*LN(F17)	=B$1*E17	=C17	=B$1*F17
12	2873	=C18+D17	=1-(G$2/(G$2+A18))^G$1		=E18-E17	=C18*LN(F18)	=B$1*E18	=C18	=B$1*F18
13	2270	=C19+D18	=1-(G$2/(G$2+A19))^G$1		=E19-E18	=C19*LN(F19)	=B$1*E19	=C19	=B$1*F19
14	3402	=C20+D19	=1-(G$2/(G$2+A20))^G$1		=E20-E19	=C20*LN(F20)	=B$1*E20	=C20	=B$1*F20
15	2461	=C21+D20	=1-(G$2/(G$2+A21))^G$1		=E21-E20	=C21*LN(F21)	=B$1*E21	=C21	=B$1*F21
16	2194	=C22+D21	=1-(G$2/(G$2+A22))^G$1		=E22-E21	=C22*LN(F22)	=B$1*E22	=C22	=B$1*F22
17	4884	=C23+D22	=1-(G$2/(G$2+A23))^G$1		=E23-E22	=C23*LN(F23)	=B$1*E23	=C23	=B$1*F23
18	2841	=C24+D23	=1-(G$2/(G$2+A24))^G$1		=E24-E23	=C24*LN(F24)	=B$1*E24	=C24	=B$1*F24
19	1365	=C25+D24	=1-(G$2/(G$2+A25))^G$1		=E25-E24	=C25*LN(F25)	=B$1*E25	=C25	=B$1*F25
20	1167	=C26+D25	=1-(G$2/(G$2+A26))^G$1		=E26-E25	=C26*LN(F26)	=B$1*E26	=C26	=B$1*F26
21	1508	=C27+D26	=1-(G$2/(G$2+A27))^G$1		=E27-E26	=C27*LN(F27)	=B$1*E27	=C27	=B$1*F27
22	2123	=C28+D27	=1-(G$2/(G$2+A28))^G$1		=E28-E27	=C28*LN(F28)	=B$1*E28	=C28	=B$1*F28
23	2010	=C29+D28	=1-(G$2/(G$2+A29))^G$1		=E29-E28	=C29*LN(F29)	=B$1*E29	=C29	=B$1*F29
24	1591	=C30+D29	=1-(G$2/(G$2+A30))^G$1		=E30-E29	=C30*LN(F30)	=B$1*E30	=C30	=B$1*F30
25	2079	=C31+D30	=1-(G$2/(G$2+A31))^G$1		=E31-E30	=C31*LN(F31)	=B$1*E31	=C31	=B$1*F31
26	1378	=C32+D31	=1-(G$2/(G$2+A32))^G$1		=E32-E31	=C32*LN(F32)	=B$1*E32	=C32	=B$1*F32
27	999	=C33+D32	=1-(G$2/(G$2+A33))^G$1		=E33-E32	=C33*LN(F33)	=B$1*E33	=C33	=B$1*F33
28	1409	=C34+D33	=1-(G$2/(G$2+A34))^G$1		=E34-E33	=C34*LN(F34)	=B$1*E34	=C34	=B$1*F34
29	1064	=C35+D34	=1-(G$2/(G$2+A35))^G$1		=E35-E34	=C35*LN(F35)	=B$1*E35	=C35	=B$1*F35

Tableau 5.8 : Paramètres du modèle Exponentiel Gamma (résultats)

Nbr	106426			r	61218510,1187				
				alpha	578521940,3574				
				LL=	-345050,8192				
								PAR JOUR	
Questions	incr_Trl	Cumul insc	P(T<=t)	P(inscription	LL		E(T(t))	Observed	Expected
1	9688	9688	0,1004	0,1004	-22267,57531		10686,49	9688,00	10686,49
2	13943	23631	0,1907	0,0903	-33522,99518		20299,92	13943,00	9613,43
3	8050	31681	0,2720	0,0813	-20206,36481		28948,05	8050,00	8648,12
4	5494	37175	0,3451	0,0731	-14371,8988		36727,79	5494,00	7779,75
5	3065	40240	0,4109	0,0658	-8342,148438		43726,35	3065,00	6998,56
6	2548	42788	0,4700	0,0592	-7204,632792		50022,17	2548,00	6295,82
7	6062	48850	0,5232	0,0532	-17782,16598		55685,82	6062,00	5663,64
8	5435	54285	0,5711	0,0479	-16518,0603		60780,76	5435,00	5094,94
9	5097	59382	0,6142	0,0431	-16030,1687		65364,11	5097,00	4583,35
10	4690	64072	0,6529	0,0387	-15246,43568		69487,23	4690,00	4123,12
11	4594	68666	0,6878	0,0349	-15420,48677		73196,34	4594,00	3709,11
12	2873	71539	0,7191	0,0314	-9947,696027		76533,01	2873,00	3336,67
13	2270	73809	0,7473	0,0282	-8100,031045		79534,63	2270,00	3001,63
14	3402	77211	0,7727	0,0254	-12499,3373		82234,86	3402,00	2700,23
15	2461	79672	0,7955	0,0228	-9302,415737		84663,95	2461,00	2429,09
16	2194	81866	0,8161	0,0205	-8525,340183		86849,13	2194,00	2185,18
17	4884	86750	0,8345	0,0185	-19494,83221		88814,89	4884,00	1965,76
18	2841	89591	0,8511	0,0166	-11640,68433		90583,26	2841,00	1768,37
19	1365	90956	0,8661	0,0149	-5737,37963		92174,07	1365,00	1590,81
20	1167	92123	0,8795	0,0134	-5028,634978		93605,14	1167,00	1431,07
21	1508	93631	0,8916	0,0121	-6657,588094		94892,51	1508,00	1287,37
22	2123	95754	0,9025	0,0109	-9597,37173		96050,62	2123,00	1158,10
23	2010	97764	0,9123	0,0098	-9299,232541		97092,43	2010,00	1041,82
24	1591	99355	0,9211	0,0088	-7529,093541		98029,64	1591,00	937,21
25	2079	101434	0,9290	0,0079	-10058,45456		98872,74	2079,00	843,10
26	1378	102812	0,9362	0,0071	-6812,749727		99631,18	1378,00	758,44
27	999	103811	0,9426	0,0064	-5044,709351		100313,46	999,00	682,28
28	1409	105220	0,9483	0,0058	-7264,209297		100927,24	1409,00	613,77
29	1064	106284	0,9535	0,0052	-5598,126132		101479,38	1064,00	552,14



Modèle Exponentiel-Gamma avec covariés :

Afin d'intégrer les variations de données observées, nous avons testé l'intégration de covariés au travers d'un modèle Exponentiel Gamma avec covariés.

Rappel sur l'intégration de covariés :

Dans le cas du modèle Exponentiel-Gamma avec covariés, on incorpore les variables covariées de la manière suivante:

- Soit x(t) le vecteur des variables covariées x_i au temps t et β les effets de ces covariés

- On fait l'hypothèse que les covariés ont un effet multiplicatif sur la base-line de la fonction de hasard suivante, définie pour la distribution exponentielle :

  $$h(t/\lambda, x(t), \beta) = \lambda \exp(\beta' x(t))$$

  Cette fonction de taux de hasard représente le taux instantané de « défaillance » en t conditionnel à la « survie » en t.

  $$h(t)=\lim_{\Delta t\to 0} P(t<T\leq t+\Delta t \mid T>t)/\Delta t=\frac{f(t)}{1-F(t)}$$

  On peut dériver la distribution correspondante (avec les covariés) en reprenant la relation fondamentale entre la fonction de hasard et la cdf d'une distribution :

  $$F(t)=1-\exp\left(\int_0^t h(u)du\right)$$

  Cette fonction F(t) est nommée régression proportionnelle de hasard

- La distribution exponentielle a un taux de hasard constant λ ($F(t)=1-e^{-\lambda t}$)

Donc en incluant les covariés, on obtient :

$$F(t) = 1 - e^{-\lambda A(t)}$$

Avec $A(t) = \Sigma^{t}_{j=1} \exp(\beta' x_i)$

- Dans le modèle exponentiel gamma, le paramètre lambda est distribué sur l'ensemble des joueurs selon une distribution gamma :

$$g(\lambda)=\frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda^r}}{\Gamma(r)}$$

où r est le paramètre de forme du modèle et α le paramètre d'échelle

- La distribution cumulée du modèle Exponentiel-Gamma avec covariés est:

$$P(T\le t)=\int_0^{\infty}(1-\exp(-\lambda A(t)))\frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}d\lambda$$

D'où $$P(T\le t)=1-\left(\frac{\alpha}{\alpha+A(t)}\right)^r$$ Equation 5.1

avec $A(t) = \Sigma^{t}_{j=1} \exp(\beta' x_i)$

Pour modéliser le nombre de joueurs issus de l'affiliation, nous avons introduit dans le modèle Exponentiel Gamma, comme variable covariée, le type de jour de jeu dans la semaine, selon l'hypothèse que celui-ci influe sur le nombre d'inscrits via l'affiliation. En effet, on peut constater dans les données observées de la figure 5.4 des pics (à la baisse) d'inscriptions tous les 6-7 jours, probablement liés au fait que les inscriptions sont moins fortes pendant le week-end (le 1er jour du jeu est un lundi). Le variable covéariée vaut donc 0 pour les jours de jeu du lundi au vendredi, et 1 pour les samedi et dimanches. Les formules utilisées sous Excel ainsi que les résultats de l'estimation des paramètres sont fournis dans les tableaux ci-dessous. La performance de ce modèle se lit sur la figure 5.8

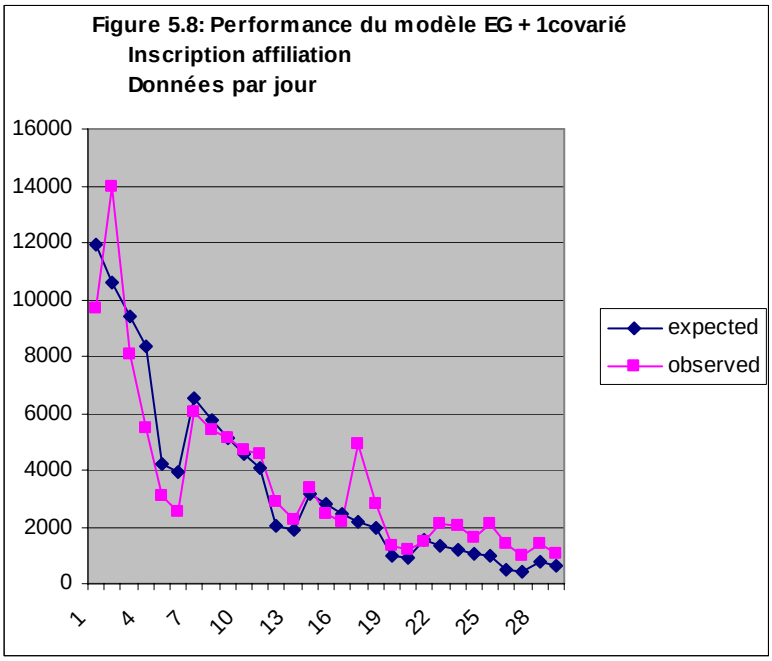


Tableau 5.9 : Paramètres du modèle Exponentiel Gamma + 1 covarié (formules)

nbre 106

r	15604,3623099
alpha	131623,690655
B_jour	-0,5965191270
LL=	=SOMME(G7:(

=TRANSPOSE(G3)

Don	Cum inscripti	Incremen	P(T<=t)	P(reponse	LL	E(T(t))	A(t)	exp(BX)	Covariates no jour
1	=D7	9688	=1-(G$2/(G$2+K7))^G$1	=E7	=D7*LN(F7)	=B$1*E7	=L7	=EXP(SOMMEPROD(L$4;N7))	0
=A7	=C7+D8	13943	=1-(G$2/(G$2+K8))^G$1	=E8-E7	=D8*LN(F8)	=B$1*E8	=K7+L8	=EXP(SOMMEPROD(L$4;N8))	0
=A8	=C8+D9	8050	=1-(G$2/(G$2+K9))^G$1	=E9-E8	=D9*LN(F9)	=B$1*E9	=K8+L9	=EXP(SOMMEPROD(L$4;N9))	0
=A9	=C9+D10	5494	=1-(G$2/(G$2+K10))^G$1	=E10-E9	=D10*LN(F10)	=B$1*E10	=K9+L10	=EXP(SOMMEPROD(L$4;N10))	0
=A1	=C10+D11	3065	=1-(G$2/(G$2+K11))^G$1	=E11-E10	=D11*LN(F11)	=B$1*E11	=K10+L11	=EXP(SOMMEPROD(L$4;N11))	1
=A1	=C11+D12	2548	=1-(G$2/(G$2+K12))^G$1	=E12-E11	=D12*LN(F12)	=B$1*E12	=K11+L12	=EXP(SOMMEPROD(L$4;N12))	1
=A1	=C12+D13	6062	=1-(G$2/(G$2+K13))^G$1	=E13-E12	=D13*LN(F13)	=B$1*E13	=K12+L13	=EXP(SOMMEPROD(L$4;N13))	0
=A1	=C13+D14	5435	=1-(G$2/(G$2+K14))^G$1	=E14-E13	=D14*LN(F14)	=B$1*E14	=K13+L14	=EXP(SOMMEPROD(L$4;N14))	0
=A1	=C14+D15	5097	=1-(G$2/(G$2+K15))^G$1	=E15-E14	=D15*LN(F15)	=B$1*E15	=K14+L15	=EXP(SOMMEPROD(L$4;N15))	0
=A1	=C15+D16	4690	=1-(G$2/(G$2+K16))^G$1	=E16-E15	=D16*LN(F16)	=B$1*E16	=K15+L16	=EXP(SOMMEPROD(L$4;N16))	0
=A1	=C16+D17	4594	=1-(G$2/(G$2+K17))^G$1	=E17-E16	=D17*LN(F17)	=B$1*E17	=K16+L17	=EXP(SOMMEPROD(L$4;N17))	0
=A1	=C17+D18	2873	=1-(G$2/(G$2+K18))^G$1	=E18-E17	=D18*LN(F18)	=B$1*E18	=K17+L18	=EXP(SOMMEPROD(L$4;N18))	1
=A1	=C18+D19	2270	=1-(G$2/(G$2+K19))^G$1	=E19-E18	=D19*LN(F19)	=B$1*E19	=K18+L19	=EXP(SOMMEPROD(L$4;N19))	1
=A1	=C19+D20	3402	=1-(G$2/(G$2+K20))^G$1	=E20-E19	=D20*LN(F20)	=B$1*E20	=K19+L20	=EXP(SOMMEPROD(L$4;N20))	0
=A2	=C20+D21	2461	=1-(G$2/(G$2+K21))^G$1	=E21-E20	=D21*LN(F21)	=B$1*E21	=K20+L21	=EXP(SOMMEPROD(L$4;N21))	0
=A2	=C21+D22	2194	=1-(G$2/(G$2+K22))^G$1	=E22-E21	=D22*LN(F22)	=B$1*E22	=K21+L22	=EXP(SOMMEPROD(L$4;N22))	0
=A2	=C22+D23	4884	=1-(G$2/(G$2+K23))^G$1	=E23-E22	=D23*LN(F23)	=B$1*E23	=K22+L23	=EXP(SOMMEPROD(L$4;N23))	0
=A2	=C23+D24	2841	=1-(G$2/(G$2+K24))^G$1	=E24-E23	=D24*LN(F24)	=B$1*E24	=K23+L24	=EXP(SOMMEPROD(L$4;N24))	0
=A2	=C24+D25	1365	=1-(G$2/(G$2+K25))^G$1	=E25-E24	=D25*LN(F25)	=B$1*E25	=K24+L25	=EXP(SOMMEPROD(L$4;N25))	1
=A2	=C25+D26	1167	=1-(G$2/(G$2+K26))^G$1	=E26-E25	=D26*LN(F26)	=B$1*E26	=K25+L26	=EXP(SOMMEPROD(L$4;N26))	1
=A2	=C26+D27	1508	=1-(G$2/(G$2+K27))^G$1	=E27-E26	=D27*LN(F27)	=B$1*E27	=K26+L27	=EXP(SOMMEPROD(L$4;N27))	0
=A2	=C27+D28	2123	=1-(G$2/(G$2+K28))^G$1	=E28-E27	=D28*LN(F28)	=B$1*E28	=K27+L28	=EXP(SOMMEPROD(L$4;N28))	0
=A2	=C28+D29	2010	=1-(G$2/(G$2+K29))^G$1	=E29-E28	=D29*LN(F29)	=B$1*E29	=K28+L29	=EXP(SOMMEPROD(L$4;N29))	0
=A2	=C29+D30	1591	=1-(G$2/(G$2+K30))^G$1	=E30-E29	=D30*LN(F30)	=B$1*E30	=K29+L30	=EXP(SOMMEPROD(L$4;N30))	0
=A3	=C30+D31	2079	=1-(G$2/(G$2+K31))^G$1	=E31-E30	=D31*LN(F31)	=B$1*E31	=K30+L31	=EXP(SOMMEPROD(L$4;N31))	0
=A3	=C31+D32	1378	=1-(G$2/(G$2+K32))^G$1	=E32-E31	=D32*LN(F32)	=B$1*E32	=K31+L32	=EXP(SOMMEPROD(L$4;N32))	1
=A3	=C32+D33	999	=1-(G$2/(G$2+K33))^G$1	=E33-E32	=D33*LN(F33)	=B$1*E33	=K32+L33	=EXP(SOMMEPROD(L$4;N33))	1
=A3	=C33+D34	1409	=1-(G$2/(G$2+K34))^G$1	=E34-E33	=D34*LN(F34)	=B$1*E34	=K33+L34	=EXP(SOMMEPROD(L$4;N34))	0
=A3	=C34+D35	1064	=1-(G$2/(G$2+K35))^G$1	=E35-E34	=D35*LN(F35)	=B$1*E35	=K34+L35	=EXP(SOMMEPROD(L$4;N35))	0

Tableau 5.10 : Paramètres du modèle Exponentiel Gamma + 1 covarié (résultats)

nbre total 106426

r	15604,3623
alpha	131623,6907
B_jour	-0,5965
LL=	-342330,6491

-5,97E-01

Données jour/heure	Cum inscripti	Incremental i	P(T<=t)	P(reponse Q	LL	E(T(t))	A(t)	exp(BX)	Covariates no jour	C
1	9688	9688	0,1118	0,1118	-21227,2893	11897,8647	1,00E+00	1,00E+00	0	
2	23631	13943	0,2111	0,0993	-32203,4570	22465,5353	2,00E+00	1,00E+00	0	
3	31681	8050	0,2993	0,0882	-19547,0837	31851,7374	3,00E+00	1,00E+00	0	
4	37175	5494	0,3776	0,0783	-13991,9373	40188,5671	4,00E+00	1,00E+00	0	
5	40240	3065	0,4170	0,0393	-9917,1711	44374,9257	4,55E+00	5,51E-01	1	
6	42788	2548	0,4538	0,0368	-8410,7203	48296,6808	5,10E+00	5,51E-01	1	
7	48850	6062	0,5149	0,0611	-16948,7895	54794,9944	6,10E+00	1,00E+00	0	
8	54285	5435	0,5691	0,0542	-15840,1018	60566,8161	7,10E+00	1,00E+00	0	
9	59382	5097	0,6173	0,0482	-15459,2836	65693,3701	8,10E+00	1,00E+00	0	
10	64072	4690	0,6601	0,0428	-14780,8603	70246,7990	9,10E+00	1,00E+00	0	
11	68666	4594	0,6981	0,0380	-15022,9386	74291,1792	1,01E+01	1,00E+00	0	
12	71539	2873	0,7171	0,0191	-11374,1181	76322,0829	1,07E+01	5,51E-01	1	
13	73809	2270	0,7350	0,0179	-9135,0657	78224,6268	1,12E+01	5,51E-01	1	
14	77211	3402	0,7646	0,0296	-11972,4897	81377,1387	1,22E+01	1,00E+00	0	
15	79672	2461	0,7909	0,0263	-8952,6254	84177,2245	1,32E+01	1,00E+00	0	
16	81866	2194	0,8143	0,0234	-8241,4283	86664,2852	1,42E+01	1,00E+00	0	
17	86750	4884	0,8351	0,0208	-18924,9922	88873,3160	1,52E+01	1,00E+00	0	
18	89591	2841	0,8535	0,0184	-11345,3709	90835,3999	1,62E+01	1,00E+00	0	
19	90956	1365	0,8628	0,0093	-6391,3203	91820,6734	1,68E+01	5,51E-01	1	
20	92123	1167	0,8714	0,0087	-5540,4164	92743,6772	1,73E+01	5,51E-01	1	
21	93631	1508	0,8858	0,0144	-6397,7805	94273,0992	1,83E+01	1,00E+00	0	
22	95754	2123	0,8986	0,0128	-9258,6237	95631,5514	1,93E+01	1,00E+00	0	
23	97764	2010	0,9099	0,0113	-9004,0905	96838,1470	2,03E+01	1,00E+00	0	
24	99355	1591	0,9200	0,0101	-7315,7190	97909,8626	2,13E+01	1,00E+00	0	
25	101434	2079	0,9289	0,0089	-9806,0826	98861,7767	2,23E+01	1,00E+00	0	
26	102812	1378	0,9334	0,0045	-7448,8745	99339,7887	2,29E+01	5,51E-01	1	
27	103811	999	0,9376	0,0042	-5465,3813	99787,5915	2,34E+01	5,51E-01	1	
28	105220	1409	0,9446	0,0070	-6996,8631	100529,6063	2,44E+01	1,00E+00	0	
29	106284	1064	0,9508	0,0062	-5409,7745	101188,6768	2,54E+01	1,00E+00	0	

On constate que le modèle EG + 1 covarié améliore nettement la vraisemblance (-342330 vs -345050) et intègre davantage les pics de variation du nombre de joueurs issus de l'affiliation. Le test du ratio de vraisemblance entre exponentiel Gamma et EG avec 1 covarié (jour de la semaine) donne LR= 5442 très nettement supérieur à 3,84 (test du Chi-2). Nous retiendrons donc pour mesurer l'affiliation un modèle Exponentiel Gamma avec 1 covarié qui est le jour de la semaine.

Les paramètres de ce modèle sont :

Tableau 5.11 :

Paramètres du modèle Exponentiel Gamma avec 1 covarié (jour de la semaine)

r	15604,3623
alpha	131623,6907
B_jour	-0,5965
LL=	-342330,6491

r et alpha sont les paramètres du modèle Gamma, B_jour est le coefficient du covarié jour de la semaine.

Durée inter inscription affiliation :

Dans les modèles ci-dessus présentés décrivant le processus des inscriptions au jeu, nous nous sommes intéressés à modéliser le nombre de joueurs par jour et à l'expliquer par un certain nombre de variables covariées, l'objectif étant de mieux comprendre le comportement des individus pour chaque action marketing initiale (affiliation, mails d'annonce ou de relance, autopromotion sur le site). Dans cette optique, nous nous sommes concentrée sur des modèles explicatifs.

Dans une deuxième étape, nous allons nous intéresser à la paramétrisation des inscriptions dans le réseau de Petri stochastique. L'objectif est ici plus simple puisque nous cherchons à modéliser, dans le réseau de Petri de la figure 5.2, la loi de la transition inscription pour chaque source de recrutement, qui est dans le réseau de Petri la loi suivie par la variable durée entre 2 inscriptions d'une même source de recrutement. Nous aurons ainsi la loi de la transition inscription affiliation (durée entre 2 inscriptions venant de l'affiliation), la loi de la transition inscription mail d'annonce (durée entre 2 inscriptions venant du mail d'annonce),

celles des transitions inscription mail de relance (durée entre 2 inscriptions venant du mail de relance), inscription autopromotion du site (durée entre 2 inscriptions venant de l'autopromotion sur le site) et inscription source non identifiée (durée entre 2 inscriptions venant de source non identifiée).

Nous allons donc ci-dessous rechercher la loi de la durée entre deux inscriptions provenant de l'affiliation.

Les statistiques principales de la variable durée entre 2 inscriptions affiliation sont fournies dans le tableau 5.12 ci-dessous :

Tableau 5.12 : Statistiques durée entre 2 inscriptions affiliation :

Statistics

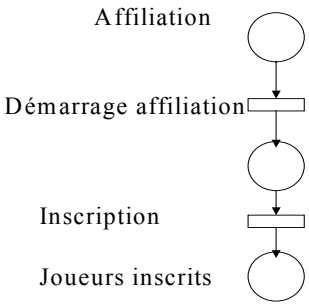
delte entre 2 inscriptions en seconde

N	Valid	106426
	Missing	1
Mean		23,2620
Median		9,0000
Std. Deviation		62,42875
Variance		3897,348
Minimum		,00
Percentiles	10	1,0000
	20	2,0000
	30	4,0000
	40	6,0000
	50	9,0000
	60	13,0000
	70	19,0000
	80	28,0000
	90	51,0000

Comme le montre le tableau 5.12 ci-dessus, la durée moyenne entre 2 inscriptions affiliation est de 23,2 secondes, plus de 90% des données étant inférieures à 51 secondes. Ceci signifie que les inscriptions via l'affiliation sont très fréquentes. La 1ère inscription étant le 10 Mai à 19h15, nous avons fait démarrer dans le réseau de Petri la 1ère inscription à cette heure-là, en rajoutant une place et une transition (démarrage inscriptions affiliation) dans le réseau permettant de prendre en compte ce délai.

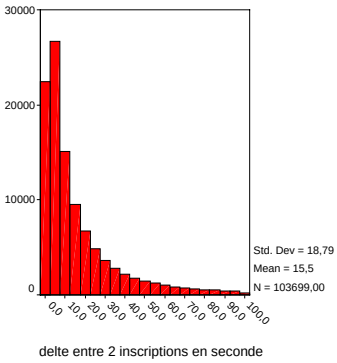
Le détail du réseau de Petri sur la partie inscription affiliation est présenté sur la figure 5.9 ci-dessous :

Figure 5.9:
Détail du réseau de Pétri: inscription affiliation



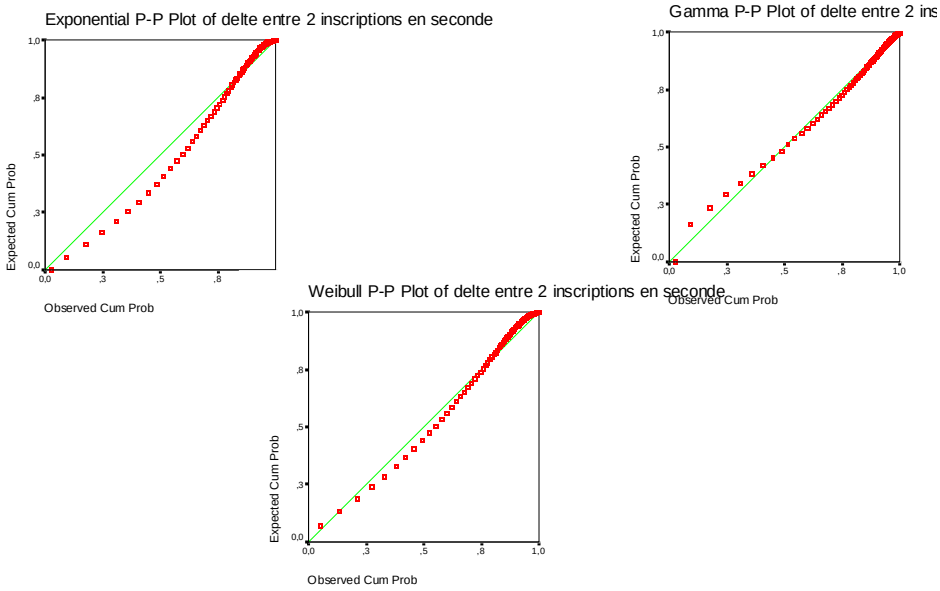
La figure 5.9 montre la transition « démarrage affiliation » qui a été rajoutée pour tenir compte du fait que le 1er jour n'est pas complet en terme d'inscriptions (la 1ère inscription démarre à 19h15). La transition « démarrage affiliation » est une loi constante retardant la première inscription de 19,25 h le premier jour.
Concernant la variable durée entre 2 inscriptions affiliation, plusieurs lois ont été testées. La figure 5.10 correspond à l'histogramme de cette variable et montre une courbe décroissante.

Figure 5.10 : Histogramme de la durée entre 2 inscriptions affiliation (en seconde) :



Nous avons donc dans un premier temps effectué des PP Plots pour trouver la loi qui approche le plus les données observées. Nous avons retenu principalement 2 lois : les lois Gamma et Weibull, les autres lois n'ayant pas donné de bons résultats. Nous présentons sur la figure 5.11 ci-dessous les PP Plots des lois exponentielle, Gamma et Weibull. La figure 5.11 montre que la loi exponentielle présente d'importants décalages entre les données observées et celles estimées, alors que les lois Gamma et Weibull représentent correctement les données observées.

Figure 5.11: PP Plots lois Exponentielle, Weibull et Gamma
Durée entre 2 inscriptions affiliation



Pour choisir entre les 2 lois Weibull et Gamma, nous avons estimé les paramètres de ces lois sous Excel en maximisant la vraisemblance. Nous obtenons une vraisemblance de –142159 pour la loi Gamma et –142157 pour la loi de Weibull, que nous avons donc retenue.
Etant donné que nous souhaitons ensuite utiliser les réseaux de Petri dans un objectif de simulation et que nous souhaitons simuler des périodes de temps différentes, nous devons valider que les paramètres de cette loi de Weibull sont stables sur l'ensemble de la période du jeu. Le tableau ci-dessous fournit la distribution des durées entre 2 inscriptions liées à l'affiliation et la durée correspondante depuis la première inscription liée à l'affiliation..

Durée moyenne depuis la première inscription	Durée Inter inscription	Nbre cas
16,4848	200+	980
16,7408	190-200	79
16,3741	180-190	95
17,282	170-180	107
16,3953	160-170	123
17,4078	150-160	142
17,1113	140-150	183
17,0395	130-140	209
17,0305	120-130	281
17,1278	110-120	325
17,2639	100-110	387
17,0885	90-100	564
16,5258	80-90	685
16,5232	70-80	978
16,1072	60-70	1356
15,4408	50-60	1860
14,4316	40-50	2845
13,5582	30-40	4548
12,4855	20-30	8156
10,3962	10-20	16821
7,1906	0-10	50997

On constate que pour la première semaine (durée = 7,19 jours), les durées entre deux inscriptions sont très courtes (moins de 10 secondes) alors qu'au delà du 15 ème jour les durées entre deux inscriptions sont plus longues (supérieures à 50 secondes). En revanche, au delà du 15ème jour, le temps entre 2 inscriptions semble moins corrélé avec la durée depuis la première inscription. Ceci est logique avec la politique d'affiliation qui est stable après (nombre de sites partenaires stable).

Afin que la loi de Weibull reflète cependant les variations notamment en début de jeu, nous proposons de faire varier l'un des paramètres de Weibull dans le temps.Les durées entre 2 inscriptions diminuant dans le temps, nous devons avoir une croissance du paramètre béta excel de la loi de Weibull.

Le modèle Weibull avec paramètre variant dans le temps est le suivant :

- Le paramètre alpha de la loi de Weibull est stable
- Le paramètre béta varie dans le temps : $\beta = \lambda_0 e^{\gamma t}$

Les résultats de ce modèle sont les suivants :

			alpha Petri		**Pétri**	
			Pétri 10j	0,213785	alpha 2	
			Pétri 20j	0,202985	béta 2	0,6000000
			Pétri 30j	0,19273		
					LOI WEIBULL EXCEL	
					alpha	0,6000000
béta excel 10 jrs	13,08290806				lambda0	12,0000000
béta excel 20 jrs	14,26354027				gamma	0,0000001
béta excel 30 jrs	15,55071549				LL=	-142142,1046880
Durée moyenne depuis le début du jeu	**Durée Inter inscription**	**Nbre cas**	% réel		P(T=t)	
16,48	200+	980	0,010684576		0,0059363	-5024,1406417
16,74	190-200	79	0,000861308		0,0002956	-642,0033846
16,37	180-190	95	0,00103575		0,0011488	-643,0551111
17,28	170-180	107	0,001166581		0,0011815	-721,2830849
16,40	160-170	123	0,001341023		0,0015306	-797,2960866
17,41	150-160	142	0,001548173		0,0017328	-902,8415692
17,11	140-150	183	0,001995181		0,0024912	-1097,0855771
17,04	130-140	209	0,002278649		0,0028134	-1227,5361619
17,03	120-130	281	0,003063639		0,0036452	-1577,6294614
17,13	110-120	325	0,003543354		0,0042042	-1778,2909076
17,26	100-110	387	0,004219317		0,0057894	-1993,7165905
17,09	90-100	564	0,006149083		0,0071826	-2783,9582833
16,53	80-90	685	0,007468301		0,0093261	-3202,3335234
16,52	70-80	978	0,010662771		0,0123148	-4300,2211310
16,11	60-70	1356	0,014783964		0,0167785	-5542,8598463
15,44	50-60	1860	0,020278889		0,0228952	-7024,8976021
14,43	40-50	2845	0,031017978		0,0330035	-9704,6997456
13,56	30-40	4548	0,049585155		0,0492985	-13688,8484705
12,49	20-30	8156	0,088921839		0,0801911	-20580,3860990
10,40	10-20	16821	0,183393116		0,1500949	-31900,8148426
7,19	0-10	50997	0,556001352		0,5881458	-27068,2065675

Ce modèle améliore significativement le niveau de vraisemblance (-142142 versus –142152) avec un seul paramètre supplémentaire et le bon niveau de R^2 (0.98) montre que ce modèle représente bien l'arrivée des joueurs par jour liée à l'affiliation. C'est par conséquent ce modèle Weibull de paramètres :

α excel = 0,60

β excel = 12 $e^{0,0000001\ t}$

La formule de la loi de Weibull est différente sur excel et dans les réseaux de Petri (le paramètre :

val1 (paramètre alpha1) dans les réseaux de Petri = (1/ Béta excel)$^{\text{alpha excel}}$)

val2 (paramètre béta1) dans les réseaux de Petri = alpha excel

En effet, les formules pour la fonction de densité de probabilité de la loi de Weibull pour chaque cas sont les suivantes :

Excel : $f(x;\alpha;\beta)=\frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}e^{-(x/\beta)^{\alpha}}$ Equation 1

Petri : $f(x;val1;val2)=val1\,val2\,x^{val2-1}e^{-val1*x^{val2}}$ Equation 2

Par conséquent, les paramètres de la loi de Weibull intégrant une variation de paramètre dans le temps sont les suivants :

$$\alpha \text{ Petri} = (0{,}083\ e^{-0{,}0000001\ t})^{0{,}60} = 0{,}2246\ e^{-0{,}00000006\ t}$$

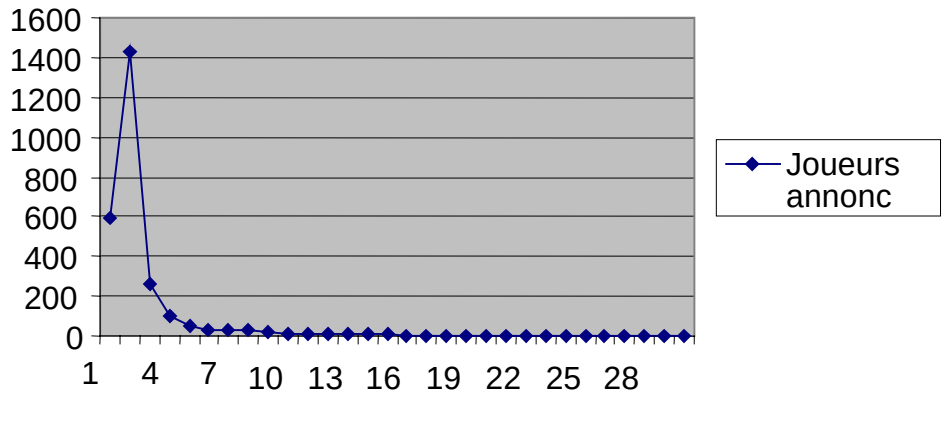
$$\beta \text{ Petri} = 0{,}60$$

Nous avons donc modélisé la durée entre deux inscriptions liées à l'affiliation. Nous allons appliquer la même démarche pour les inscriptions liées au mail d'annonce.

5.4.1.2 Modélisation du nombre d'inscriptions venant du mail d'annonce

Nous cherchons dans cette partie à modéliser le nombre de joueurs recrutés par le mail d'annonce envoyé à la base de clients Anyway le premier jour du jeu.
Comme le montre la figure 5.14 ci-dessous représentant les données observées du nombre de joueurs issus du mail annonce pour chaque jour du jeu, le nombre d'inscrits provenant du mail d'annonce varie d'un jour à l'autre avec un pic au 2ème jour du jeu.

Figure 5.14: Evolution du nombre de joueurs issus du mail annonce



On constate sur la figure 5.14 la même faible valeur du nombre d'inscriptions le premier jour, probablement dûe à l'heure d'envoi de l'e-mail d'annonce. En effet, si l'e-mail a été envoyé

dans l'après midi ou en fin de journée, cela explique le faible nombre d'inscriptions du premier jour (la première inscription liée au mail d'annonce a été enregistrée le premier jour à 18h52). Ceci s'ajoute également au phénomène selon lequel l'e-mail est effectivement envoyé le premier jour mais ouvert par certains internautes seulement le 2ème jour ou les jours suivants. En raison de ces hypothèses, nous avons pour le mail d'annonce enlevé le premier jour du jeu dans nos données (jour considéré comme le jour d'envoi de l'e-mail d'annonce) afin d'estimer les paramètres des modèles que nous avons appliqués.

Comme pour l'affiliation, nous avons testé plusieurs modèles de type exponentiel : un modèle exponentiel, un modèle Exponentiel Gamma introduisant l'hétérogénéité entre les indivbdus, un modèle Exponentiel Gamma avec covariés que nous détaillerons ultérieurement.

Les paramètres ont été estimés par la méthode du maximum de vraisemblance sous Excel. Les formules utilisées sous Excel sont les mêmes que celles utilisées pour l'affiliation (tableaux 5.5, 5.7, 5.9). Par conséquent, pour simplifier la présentation, seuls les tableaux des résultats (et non pas les formules) seront fournis dans cette partie.

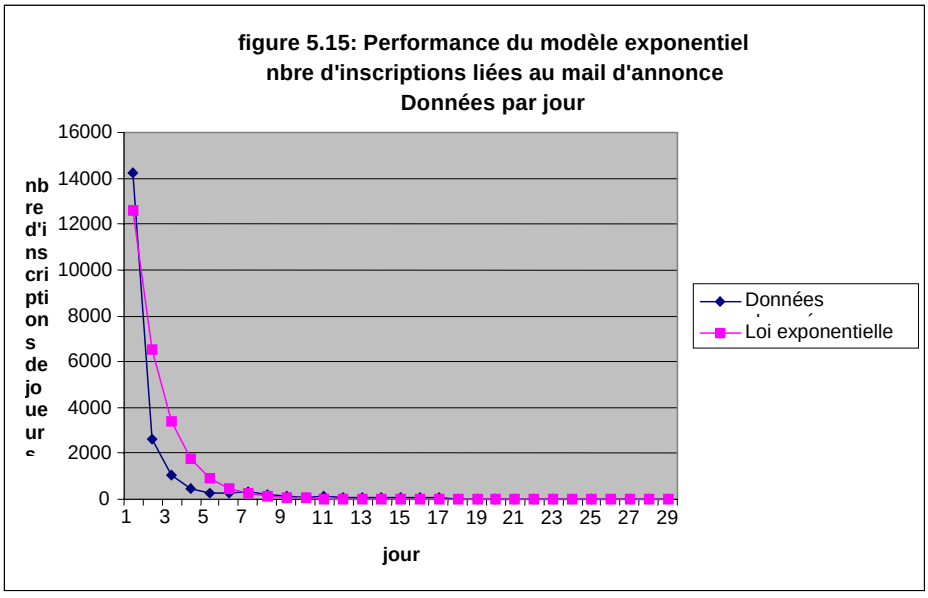
Modèle exponentiel :

Le modèle exponentiel appliqué aux données du mail d'annonce donne les résultats suivants :

Tableau 5.14 : Paramètres de la loi exponentielle calculés sous Excel (Résultats)

Nbre de joueurs	26110			lambda	0,660762			
				LL=	-28933,6			
						CUMUL	PAR JOUR	
	Jour	Cum incscriptions	Incremental	P(T<=t)	LL	E(T(t))	P(T=t)	Expected
1	11/5/04	14255	14255	0,483543	-10357,9	12625,2955	0,4835	12625,2955
2	12/5/04	16848	2593	0,733272	-3597,47	19145,7237	0,2497	6520,4282
3	13/5/04	17901	1053	0,862246	-2156,69	22513,2475	0,1290	3367,5238
4	14/5/04	18387	486	0,928856	-1316,53	24252,4303	0,0666	1739,1828
5	15/5/04	18675	288	0,963257	-970,464	25150,6443	0,0344	898,2140
6	16/5/04	18928	253	0,981024	-1019,7	25614,5336	0,0178	463,8893
7	17/5/04	19272	344	0,9902	-1613,77	25854,1127	0,0092	239,5791
8	18/5/04	19451	179	0,994939	-957,999	25977,8451	0,0047	123,7324
9	19/5/04	19579	128	0,997386	-769,627	26041,7476	0,0024	63,9025
10	20/5/04	19654	75	0,99865	-500,511	26074,7505	0,0013	33,0029
11	21/5/04	19753	99	0,999303	-726,09	26091,7952	0,0007	17,0446
12	22/5/04	19805	52	0,99964	-415,74	26100,5980	0,0003	8,8028
13	23/5/04	19871	66	0,999814	-571,28	26105,1443	0,0002	4,5463
14	24/5/04	19950	79	0,999904	-736,006	26107,4922	0,0001	2,3480
15	25/5/04	19997	47	0,99995	-468,933	26108,7048	0,0000	1,2126
16	26/5/04	20033	36	0,999974	-382,97	26109,3311	0,0000	0,6263
17	27/5/04	20073	40	0,999987	-451,953	26109,6545	0,0000	0,3234
18	28/5/04	20088	15	0,999993	-179,394	26109,8216	0,0000	0,1670
19	29/5/04	20100	12	0,999996	-151,444	26109,9079	0,0000	0,0863
20	30/5/04	20113	13	0,999998	-172,654	26109,9524	0,0000	0,0446
21	31/5/04	20125	12	0,999999	-167,302	26109,9754	0,0000	0,0230
22	1/6/04	20137	12	1	-175,232	26109,9873	0,0000	0,0119
23	2/6/04	20147	10	1	-152,634	26109,9934	0,0000	0,0061
24	3/6/04	20161	14	1	-222,938	26109,9966	0,0000	0,0032
25	4/6/04	20168	7	1	-116,094	26109,9983	0,0000	0,0016
26	5/6/04	20175	7	1	-120,72	26109,9991	0,0000	0,0008
27	6/6/04	20184	9	1	-161,158	26109,9995	0,0000	0,0004
28	7/6/04	20195	11	1	-204,239	26109,9998	0,0000	0,0002
29	8/6/04	20200	5	1	-96,1398	26109,9999	0,0000	0,0001

La performance du modèle exponentiel par rapport aux données observées est présentée sous forme graphique par la figure 5.15 ci-dessous. On constate globalement une approximation des données par le modèle mais avec des écarts, notamment sur la première semaine du jeu pour laquelle la loi exponentielle surestime le nombre d'inscriptions par mail d'annonce.



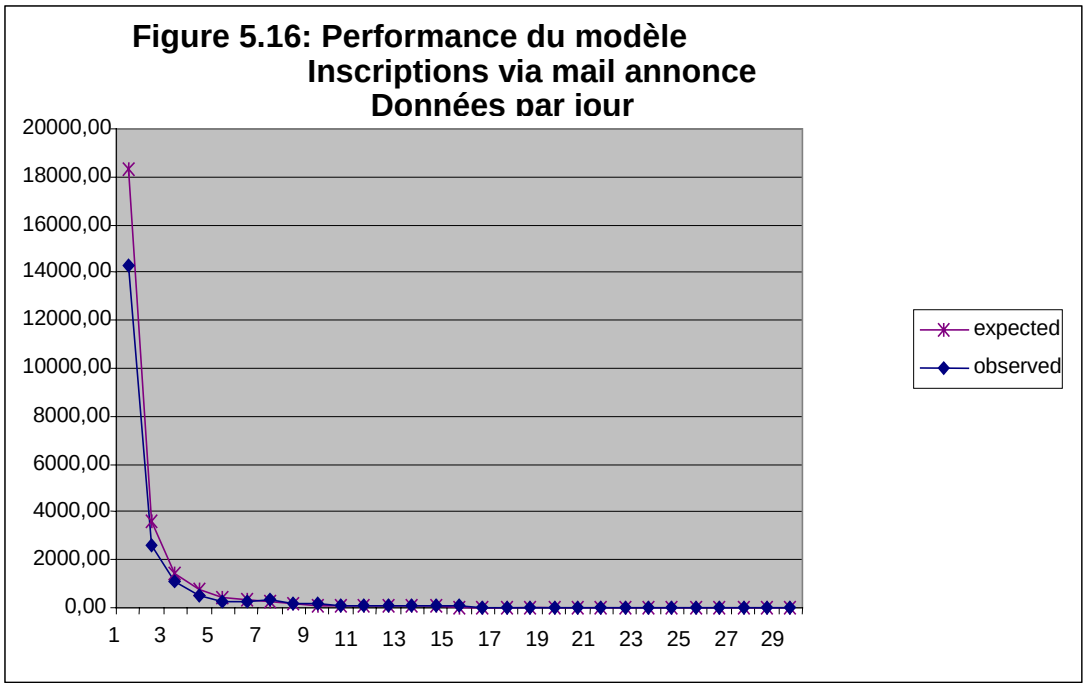
Ces différences en début de jeu peuvent provenir de plusieurs sources. L'une des explications peut être l'hétérogénéité entre les joueurs qui n'est pas intégrée dans le modèle exponentiel. Une autre explication peut être un impact différent du mail d'annonce selon les jours qui suivent l'envoi de ce mail (impact plus fort les jours qui suivent le mail d'annonce et moins fort les jours qui suivent).

Modèle Exponentiel-Gamma :

Pour améliorer le modèle, nous avons donc introduit une hétérogénéité entre les joueurs et testé un modèle EG, le paramètre lambda de la loi exponentielle suivant une loi Gamma.
Les résultats de ce modèle sont présentés dans le tableau 5.15 ci-dessous.

Tableau 5.15 : Paramètres du modèle Exponentiel Gamma (résultats)

Nbre joueurs	26110			r	1,3141				
				alpha	0,6646				
				LL=	-24590,5497				
								PAR JOUR	
Jour	Nbre c	incr_Tr	Cumul inscripsi	P(T<=t)	P(inscription)	LL	E(T(t))	Observed	Expected
1	14255	14255	0,7007	0,7007	-5069,2023	18296,50	14255,00	18296,50	
2	2593	16848	0,8387	0,1380	-5135,7520	21899,25	2593,00	3602,75	
3	1053	17901	0,8939	0,0552	-3050,8141	23339,84	1053,00	1440,60	
4	486	18387	0,9227	0,0288	-1723,5675	24092,52	486,00	752,68	
5	288	18675	0,9401	0,0174	-1166,6751	24546,99	288,00	454,46	
6	253	18928	0,9517	0,0115	-1129,4250	24847,64	253,00	300,65	
7	344	19272	0,9598	0,0081	-1656,0827	25059,49	344,00	211,85	
8	179	19451	0,9658	0,0060	-916,1174	25215,84	179,00	156,35	
9	128	19579	0,9703	0,0046	-689,4505	25335,39	128,00	119,55	
10	75	19654	0,9739	0,0036	-422,0021	25429,40	75,00	94,01	
11	99	19753	0,9768	0,0029	-578,5925	25505,01	99,00	75,62	
12	52	19805	0,9792	0,0024	-314,2504	25566,99	52,00	61,98	
13	66	19871	0,9812	0,0020	-410,9428	25618,60	66,00	51,61	
14	79	19950	0,9828	0,0017	-505,2902	25662,15	79,00	43,55	
15	47	19997	0,9843	0,0014	-308,0446	25699,34	47,00	37,19	
16	36	20033	0,9855	0,0012	-241,2750	25731,41	36,00	32,07	
17	40	20073	0,9866	0,0011	-273,6449	25759,32	40,00	27,91	
18	15	20088	0,9875	0,0009	-104,5841	25783,80	15,00	24,48	
19	12	20100	0,9883	0,0008	-85,1566	25805,42	12,00	21,62	
20	13	20113	0,9891	0,0007	-93,7842	25824,64	13,00	19,22	
21	12	20125	0,9897	0,0007	-87,9149	25841,82	12,00	17,18	
22	12	20137	0,9903	0,0006	-89,1977	25857,26	12,00	15,44	
23	10	20147	0,9909	0,0005	-75,3531	25871,20	10,00	13,94	
24	14	20161	0,9913	0,0005	-106,8642	25883,84	14,00	12,64	
25	7	20168	0,9918	0,0004	-54,0892	25895,35	7,00	11,51	
26	7	20175	0,9922	0,0004	-54,7207	25905,87	7,00	10,52	
27	9	20184	0,9926	0,0004	-71,1367	25915,51	9,00	9,64	
28	11	20195	0,9929	0,0003	-87,8654	25924,37	11,00	8,87	
29	11	20206	0,9932	0,0003	-88,7538	25932,55	11,00	8,18	



On constate que le modèle Exponentiel Gamma améliore la vraisemblance (-24590 versus –28933 pour le modèle exponentiel) et représente mieux les données. Cependant, on constate des écarts sur les premiers jours (sur estimation par le modèle). Nous avons donc introduit des covariés pour améliorer ce modèle.

Modèle Exponentiel-Gamma avec covariés :

La première variable covariée que nous avons introduite est le jour de la semaine. Cette variable sera introduite en binaire comme covariée : elle vaut 1 pour les samedis et dimanches et 0 pour les autres jours de la semaine, comme nous l'avons déjà effectué dans le cadre de l'affiliation.

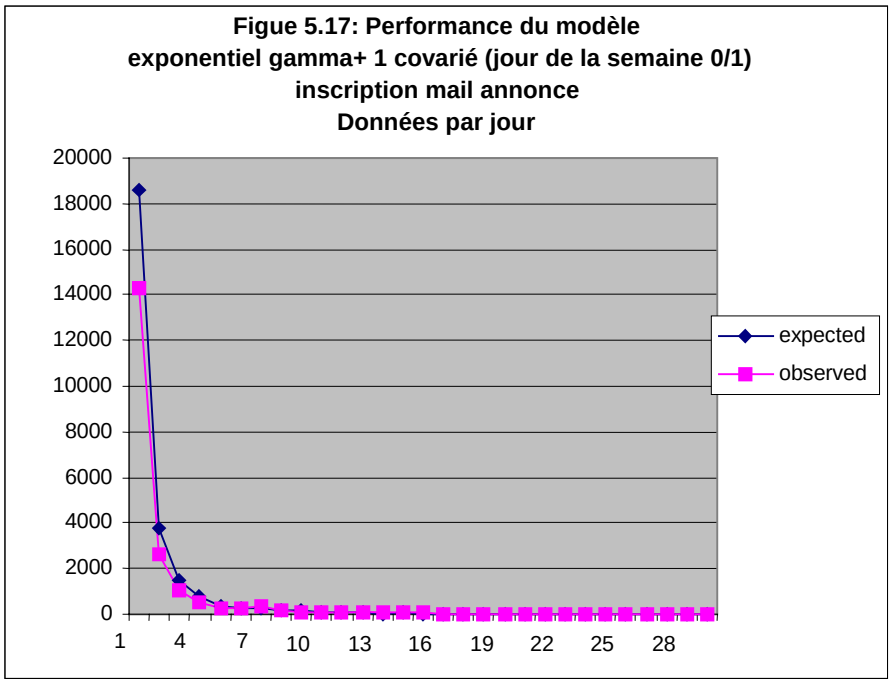
Tableau 5.16: Paramètres du modèle Exponentiel Gamma + 1 covarié (résultats)

nbre tot 26610

r	1,3935
alpha	0,7296
B_joursemaine	-0,3951
LL=	-24546,3613

-0,395072

Donnée	Jour	Cum in	Incremen	P(T<=t)	P(reponse Question	LL	E(T(t))	A(t)	exp(BX)	Covariates jour
1	11-mai	14255	14255	0,6996	0,6996	-5091,6275	18617,5599	1,0000	1	0
2	12-mai	16848	2593	0,8410	0,1413	-5073,8838	22377,9590	2,0000	1	0
3	13-mai	17901	1053	0,8971	0,0561	-3033,3484	23870,7005	3,0000	1	0
4	14-mai	18387	486	0,9261	0,0290	-1720,4964	24642,6556	4,0000	1	0
5	15-mai	18675	288	0,9386	0,0125	-1261,5373	24975,8427	4,6736	0,673632	1
6	16-mai	18928	253	0,9479	0,0093	-1184,1558	25222,6443	5,3473	0,673632	1
7	17-mai	19272	344	0,9578	0,0100	-1585,1326	25488,0074	6,3473	1	0
8	18-mai	19451	179	0,9649	0,0071	-885,8111	25676,7486	7,3473	1	0
9	19-mai	19579	128	0,9702	0,0053	-671,5761	25816,8483	8,3473	1	0
10	20-mai	19654	75	0,9742	0,0040	-413,3673	25924,3470	9,3473	1	0
11	21-mai	19753	99	0,9774	0,0032	-569,2454	26009,0447	10,3473	1	0
12	22-mai	19805	52	0,9792	0,0018	-329,1146	26056,5061	11,0209	0,673632	1
13	23-mai	19871	66	0,9808	0,0016	-426,7895	26097,8752	11,6945	0,673632	1
14	24-mai	19950	79	0,9827	0,0020	-492,2239	26150,2466	12,6945	1	0
15	25-mai	19997	47	0,9844	0,0016	-301,2393	26194,0493	13,6945	1	0
16	26-mai	20033	36	0,9858	0,0014	-236,7208	26231,1436	14,6945	1	0
17	27-mai	20073	40	0,9870	0,0012	-269,2400	26262,8984	15,6945	1	0
18	28-mai	20088	15	0,9880	0,0010	-103,1540	26290,3415	16,6945	1	0
19	29-mai	20100	12	0,9886	0,0006	-88,6586	26306,7998	17,3682	0,673632	1
20	30-mai	20113	13	0,9892	0,0006	-97,2055	26321,8545	18,0418	0,673632	1
21	31-mai	20125	12	0,9899	0,0008	-86,2569	26341,9594	19,0418	1	0
22	01-juin	20137	12	0,9906	0,0007	-87,7110	26359,7700	20,0418	1	0
23	02-juin	20147	10	0,9912	0,0006	-74,2458	26375,6405	21,0418	1	0
24	03-juin	20161	14	0,9917	0,0005	-105,4845	26389,8575	22,0418	1	0
25	04-juin	20168	7	0,9922	0,0005	-53,4785	26402,6550	23,0418	1	0
26	05-juin	20175	7	0,9925	0,0003	-56,8385	26410,5740	23,7154	0,673632	1
27	06-juin	20184	9	0,9928	0,0003	-73,6718	26417,9874	24,3891	0,673632	1
28	07-juin	20195	11	0,9932	0,0004	-86,5690	26428,1543	25,3891	1	0
29	08-juin	20206	11	0,9935	0,0003	-87,5776	26437,4304	26,3891	1	0



On constate que le modèle EG + 1 covarié améliore légèrement la vraisemblance (-24546 vs –24590 pour le modèle EG) et intègre mieux les variations d'inscriptions observés au début du jeu. Cependant, il reste toujours un écart entre données observées et estimées par le modèle pour le 2ème jour du jeu, qui est la première observation ici puisqu'on a enlevé le premier jour du jeu.

Le test du ratio de vraisemblance permettant de comparer les modèles entre eux donne les résultats suivants :

Tableau 5.17: Comparaison des modèles

			Comparaison du modèle vs modèle ligne précédente	
	LL	Nbre de paramètres	LR	Khi 2 (0,05;1)
Exponentiel	-28933,00	1		
EG	-24590,00	2	8686	3,841
EG+ 1 covarié	-24546,00	3	88	3,841

Le tableau 5.17 nous permet de conclure que :

- le modèle EG+ 1 covariés est meilleur que le modèle EG (88 > 3,841) mais l'apport du covarié est faible.
- le modèle EG est meilleur que le modèle Exponentiel (8686 > 3,841)

On en conclut que le meilleur modèle est le modèle Exponentiel Gamma avec 1 covarié (Type de jour paramétré 0 pour les jours de lundi à vendredi et 1 pour les samedis et dimanches).

Rappelons ici l'équation du modèle Exponentiel Gamma avec covariés déjà présentée précédemment (équation 5.1) :

$$P(T \leq t) = 1 - \left(\frac{\alpha}{\alpha + A(t)} \right)^{r} \qquad \text{Equation 5.1}$$

avec $A(t) = \Sigma^{t}_{j=1} \exp(\beta' x_i)$

Les paramètres du modèle retenu pour le nombre d'inscriptions issu du mail d'annonce sont:

Tableau 5.18: Paramètres du modèle EG + 1 Covarié (type de jour de la semaine)

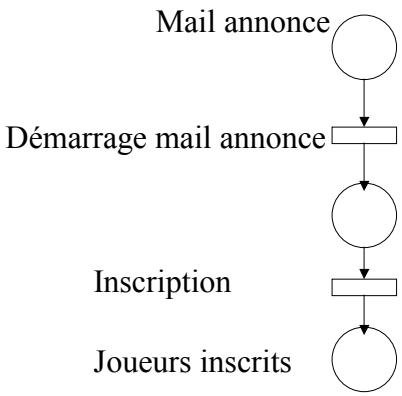
r	1,3935
alpha	0,7296
B_joursemaine	-0,3951

r et alpha sont les paramètres de la loi Gamma suivie par le paramètre lambda de la loi exponentielle. B_joursemaine est le coefficient β du covarié type de jour dans la semaine. Selon l'équation 5.1 (équation de la loi Exponentielle Gamma avec covariés), on peut en conclure que la probabilité d'arrêt les samedis et dimanches est plus faible que pour les autres jours de la semaine.

Durée inter inscription mail d'annonce :

Comme nous l'avons fait pour l'affiliation, nous allons dans une deuxième étape nous intéresser à la loi de la durée entre 2 inscriptions liées au mail d'annonce, afin d'intégrer cette loi dans le réseau de Petri de la figure 5.2 (transition inscription liée au mail de relance). Nous cherchons donc dans cette partie à modéliser la durée entre 2 inscriptions issues du mail d'annonce de l'opération envoyé à des internautes de la base de données de clients d'Anyway. La première inscription liée au mail d'annonce s'est faite le 10 mai (1er jour du jeu) à 18h52. Nous avons donc inséré dans le réseau de Petri une transition (démarrage inscriptions mailannonce) et une place supplémentaire permettant de tenir compte de ce délai de démarrage (selon le même processus que pour l'affiliation), comme le montre la figure 5.18.

Figure 5.18:
Détail du réseau de Pétri: inscription mail annonce



Les principales statistiques de la variable durée entre 2 inscriptions mail d'annonce sont fournies dans le tableau 5.19 ci-dessous.

Statistics

delta time entre 2 inscriptions en seconde

N	Valid	26110
	Missing	1
Mean		95,8384
Median		4,0000
Std. Deviation		1012,912
Variance		1025990
Minimum		,00
Percentiles	10	,0000
	20	1,0000
	30	2,0000
	40	2,0000
	50	4,0000
	60	5,0000
	70	9,0000
	80	18,0000
	90	64,0000

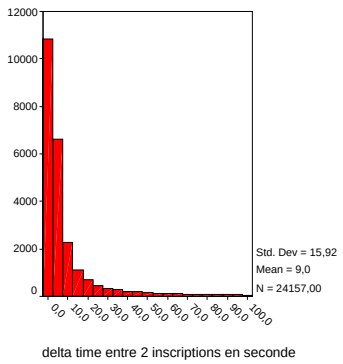
Tableau 5.19 : Statistiques de la variable durée entre 2 inscriptions par mail d'annonce

Le tableau 5.19 montre qu'en moyenne, il y a une inscription liée au mail d'annonce toutes les 95,8 secondes, mais cette moyenne reflètent des cas extrêmes (cf le maximum de 1025990

secondes) puisque 90% des données sont inférieures à 64 secondes et la médiane est égale à 4 secondes.

La figure 5.19 montre un histogramme décroissant avec une forte pente pour les premières tranches de données (pour des raisons de visibilité, l'histogramme représente 98% des données, les 2% de données extrêmes (durée entre 2 inscriptions proches du maximum) ayant été enlevées pour ce graphe, mais gardées pour l'estimation des paramètres).

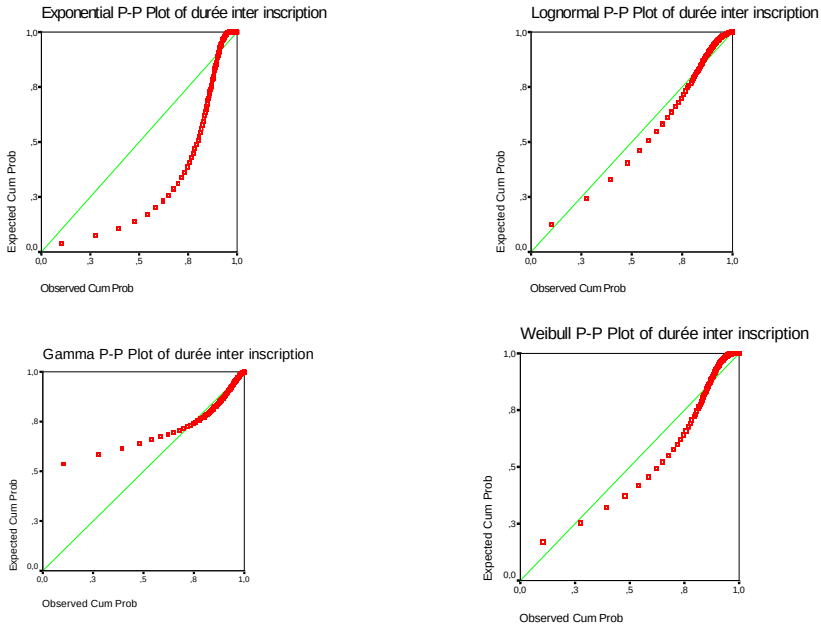
Figure 5.19 : Histogramme de la durée entre 2 inscriptions
mail annonce
(en seconde) :



Nous avons donc procédé comme pour l'affiliation et avons testé, tout d'abord par PP Plot, plusieurs lois. Les résultats sont fournis dans la figure 5.20.

Figure 5.20 : PP Plots des lois Exponentielle, Log-normale, Gamma et Weibull

Durée entre 2 inscriptions mail annonce



L'étude de ces PP Plots montre clairement que les lois Exponentielles et Gamma ne peuvent pas être retenues en raison du fort décalage entre les données et le modèle. En revanche, les lois Weibull et Log-Normale donnent une bonne représentation du phénomène observée.
Afin de choisir entre ces 2 lois, nous avons estimé sous excel les paramètres de ces 2 modèles en maximisant la vraisemblance.
La vraisemblance obtenue pour la loi de Weibull est de –29345 contre –33584 pour la loi log-normale.

Etant donné que nous souhaitons ensuite utiliser les réseaux de Petri dans un objectif de simulation et que nous souhaitons simuler des périodes de temps différentes, nous devons valider que les paramètres de cette loi de Weibull sont stables sur l'ensemble de la période du jeu. Le tableau ci-dessous fournit la distribution des durées entre 2 inscriptions liées au mail annonce et la durée correspondante depuis le début du jeu. On constate que pour les premiers jours de jeu les durées entre deux inscriptions sont très courtes (moins de 10 secondes) alors qu'au delà de 10 jours de jeu les durées entre deux inscriptions sont plus longues (supérieures à 200 secondes). Ceci est logique avec le principe de l'e-mail d'annonce dont l'impact sur les

inscriptions est fort pour les jours suivant la réception de l'e-mail et de moins en moins fort les jours suivants.

Durée moyenne depuis le début du jeu	Durée Inter inscription	Nbre cas
10,34	200+	1260
6,24	190-200	57
6,52	180-190	51
5,83	170-180	52
5,59	160-170	66
4,89	150-160	65
5,40	140-150	62
5,15	130-140	68
5,26	120-130	82
4,59	110-120	99
4,78	100-110	107
4,57	90-100	136
4,35	80-90	155
4,12	70-80	190
4,04	60-70	203
3,67	50-60	280
3,44	40-50	381
3,00	30-40	549
2,62	20-30	879
2,01	10-20	2159
1,39	0-10	14955

Afin que la loi de Weibull reflète ces variations, nous proposons de faire varier l'un des paramètres de Weibull dans le temps.Les durées entre 2 inscriptions diminuant dans le temps, nous devons avoir une croissance du paramètre béta excel de la loi de Weibull.
Le modèle Weibull avec paramètre variant dans le temps est le suivant :

- Le paramètre alpha de la loi de Weibull est stable
- Le paramètre béta varie dans le temps : $\beta = \lambda_0 e^{\gamma t}$

Les résultats de ce modèle sont les suivants :

			alpha Petri		**Pétri**	
			Pétri 10j	0,545129	alpha 2	
			Pétri 20j	0,530722	béta 2	0,3100000
			Pétri 30j	0,516696		
					LOI WEIBULL EXCEL	
					alpha	0,3100000
béta excel 10 jrs	7,079495702				lambda0	6,4935065
béta excel 20 jrs	7,718365946				gamma	0,0000001
béta excel 30 jrs	8,414889334				LL=	-29239,0826325
Durée moyenne depuis le début du jeu	Durée Interinscription	Nbre cas	% réel		P(T=t)	
10,34	200+	1260	0,057650073		0,0599250	-3546,4725793
6,24	190-200	57	0,00260798		0,0008336	-404,1165349
6,52	180-190	51	0,002333455		0,0030265	-295,8186036
5,83	170-180	52	0,002379209		0,0028248	-305,2049173
5,59	160-170	66	0,003019766		0,0033288	-376,5398669
4,89	150-160	65	0,002974012		0,0034230	-369,0202128
5,40	140-150	62	0,00283675		0,0044402	-335,8575321
5,15	130-140	68	0,003111274		0,0045079	-367,3311488
5,26	120-130	82	0,00375183		0,0052583	-430,3320636
4,59	110-120	99	0,004529649		0,0054672	-515,6900259
4,78	100-110	107	0,004895681		0,0067749	-534,4149497
4,57	90-100	136	0,006222548		0,0075409	-664,6887682
4,35	80-90	155	0,007091874		0,0087949	-733,7053439
4,12	70-80	190	0,008693265		0,0104287	-867,0059770
4,04	60-70	203	0,009288067		0,0127570	-885,4201780
3,67	50-60	280	0,012811127		0,0156577	-1163,9026797
3,44	40-50	381	0,017432284		0,0203789	-1483,3302086
3,00	30-40	549	0,02511896		0,0276976	-1968,9391270
2,62	20-30	879	0,040217789		0,0417241	-2792,2977803
2,01	10-20	2159	0,098782943		0,0759625	-5564,8549565
1,39	0-10	14955	0,684251464		0,6792476	-5784,1391785

Ce modèle améliore significativement le niveau de vraisemblance (-29239 versus –29352) avec un seul paramètre supplémentaire et le bon niveau de R^2 (0.98) montre que ce modèle représente bien l'arrivée des joueurs par jour liée au mail d'annonce. C'est par conséquent ce modèle Weibull de paramètres :

$$\alpha \text{ excel} = 0,31$$

$$\beta \text{ excel} = 6,4935\ e^{0,0000001\ t}$$

Les paramètres adaptés aux réseaux de Petri sont par conséquent :

$$\alpha \text{ Petri} = (0,154\ e^{-0,0000001\ t})^{0,31} = 0,56\ e^{-0,000000031\ t}$$

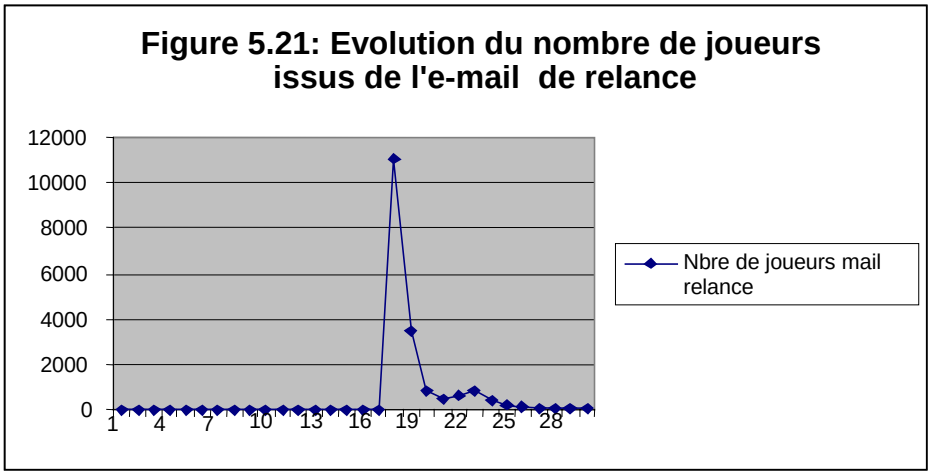
$$\beta \text{ Petri} = 0,31$$

(t étant exprimé en secondes)

Nous avons donc modélisé la durée entre deux inscriptions liées au mail d'annonce. Nous allons appliquer la même démarche pour les inscriptions liées au mail de relance.

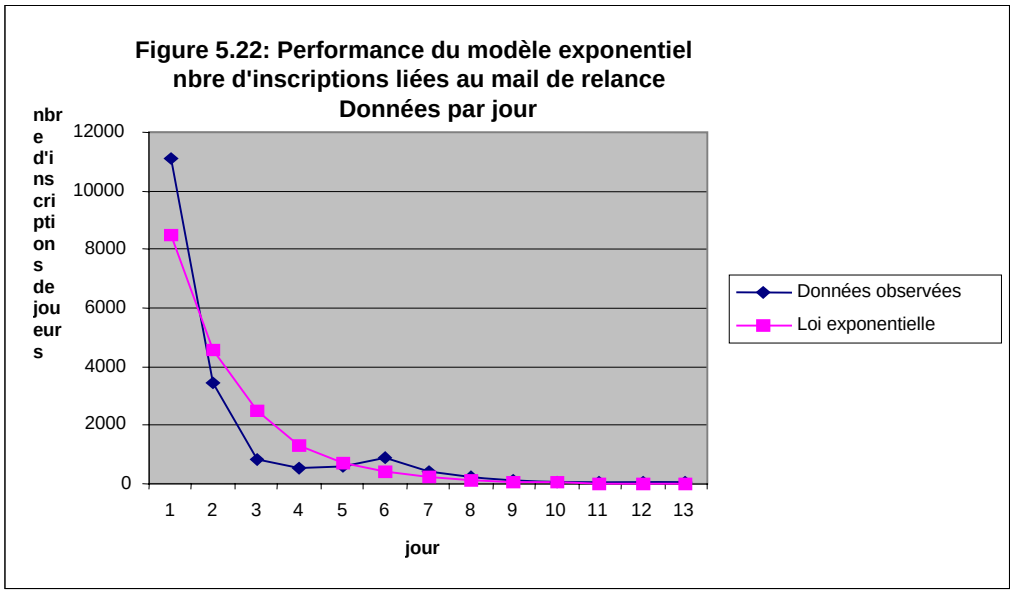
5.4.1.3 Modélisation du nombre d'inscriptions venant du mail de relance

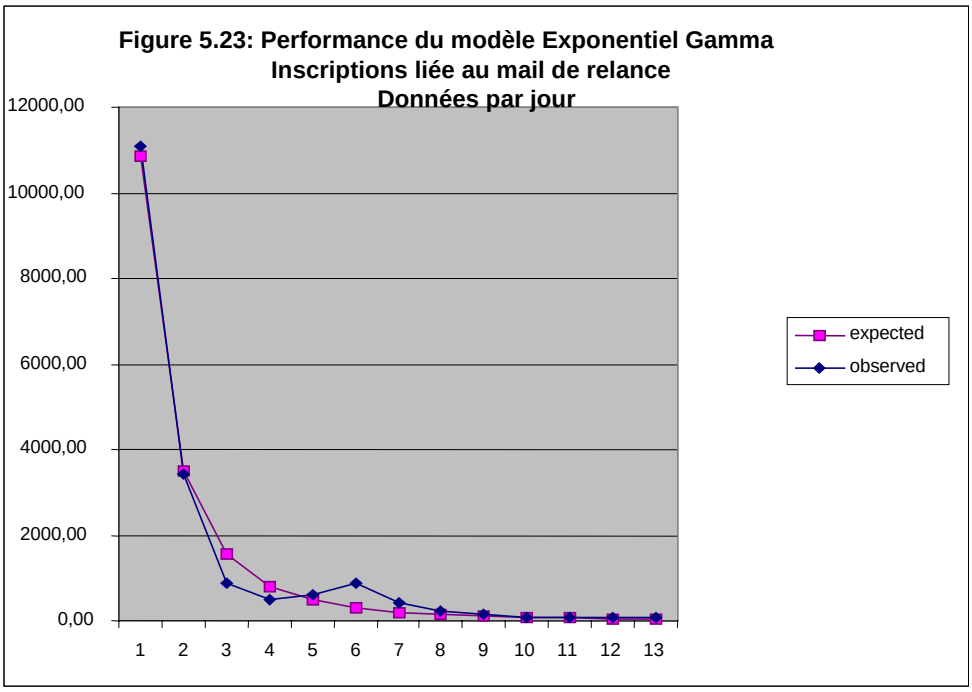
Les inscriptions au mail de relance, comme le montre la figure 5.21 ci-dessous, présentent la particularité qui est de démarrer au moment du mail de relance, soit à partir du 18ème jour de jeu. En effet, ce jour-là, la société Millermercis a envoyé un mail de relance sur les internautes de la base Anyway qui avaient reçu le mail d'annonce du jeu mais qui n'avaient pas joué.
Nous ne disposons pas dans nos données de l'heure à laquelle cet e-mail a été envoyé. En revanche, la première inscription liée à ce mail de relance a été enregistrée le 18ème jour à 12h22.



Nous modéliserons donc le nombre d'inscriptions liées au mail de relance à partir de ce 18ème jour, puisqu'il n'y a aucune inscription liée à cette action auparavant. Nous avons donc 13 jours de données pour calculer les paramètres du modèle.
Pour modéliser les inscriptions liées au mail de relance, nous avons procédé de la même manière que pour l'affiliation ou le mail d'annonce.
Nous avons donc testé un modèle exponentiel, un modèle exponentiel gamma, un modèle EG + 1 covarié (relance par mail, variable de type 1/0), un modèle EG+ 2 Covariés (relance par mail + type de jour de la semaine(samedis/dimanches versus autres jours).

Les résultats des performances de ces différents modèles sont présentés dans les figures ci-dessous :





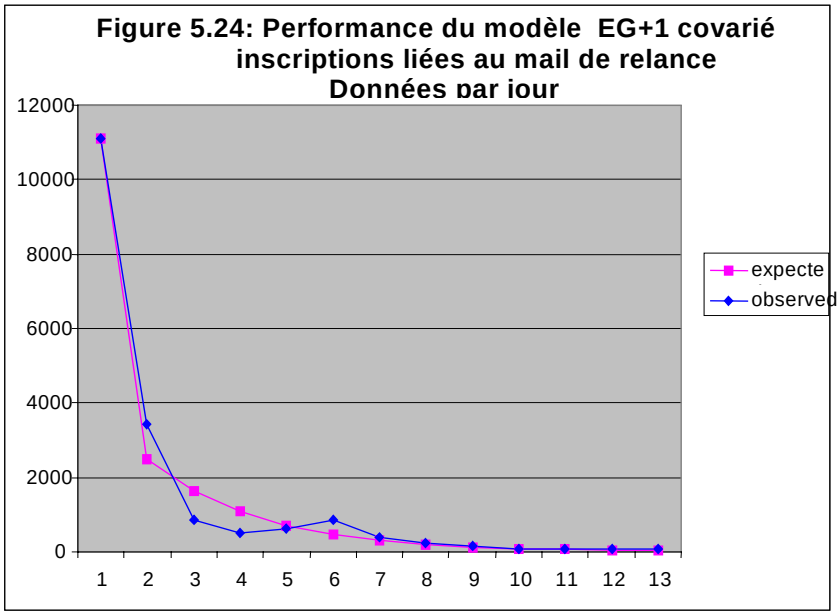


Figure 5.24: Performance du modèle EG+1 covarié
inscriptions liées au mail de relance
Données par jour
12000
10000
8000
6000
4000
2000
0
1 2 3 4 5 6 7 8 9 10 11 12 13
expecte
observed

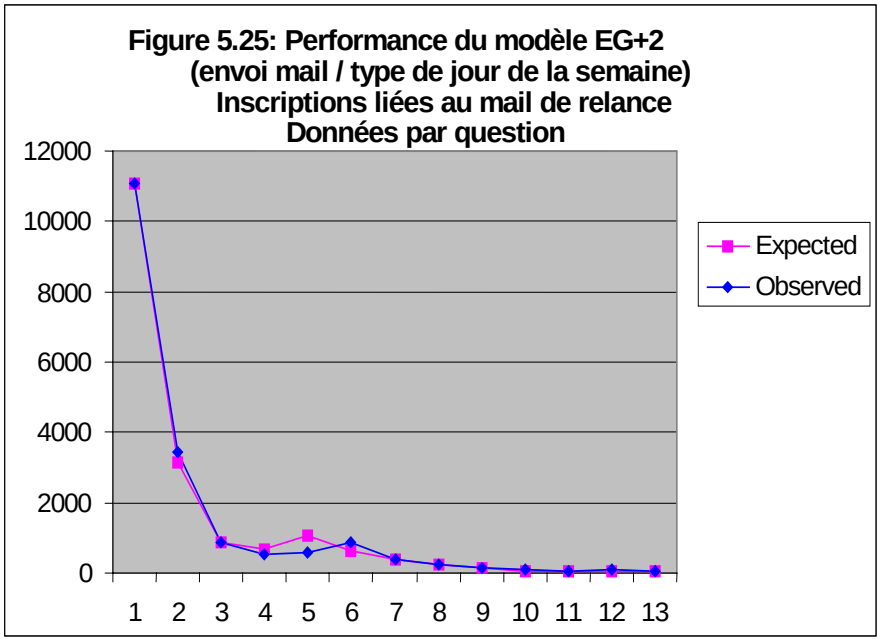


Figure 5.25: Performance du modèle EG+2
(envoi mail / type de jour de la semaine)
Inscriptions liées au mail de relance
Données par question
12000
10000
8000
6000
4000
2000
0
1 2 3 4 5 6 7 8 9 10 11 12 13
Expected
Observed

On constate que les modèles sans covariés exponentiel et exponentiel gamma ne reflètent pas les variations d'inscriptions par jour, en particulier les 4 jours qui suivent le mail de relance. En revanche, lorsqu'on introduit en covariés le type du jour de la semaine (variable paramétrée 0 pour les jours du lundi à vendredi et 1 pour les samedis et dimanches) et le fait d'avoir envoyé l'e-mail de relance (variable égale à 1 sur uniquement le 18ème jour du jeu qui correspond au jour 1 des données modélisées pour l'e-mail de relance), on améliore nettement les performances du modèle.

L'analyse du test du ratio de vraisemblance pour comparer les modèles donne les résultats suivants :

Tableau 5.22: Comparaison des modèles

			Comparaison du modèle vs modèle ligne précédente	
	LL	Nbre de paramètres	LR	Khi 2 (0,05;1)
Exponentiel	-27626,48	1		
EG	-26397,04	2	2459	3,841
EG+ 1 covarié	-26242,04	3	310	3,841
EG+2covariés	-25739,39	4	1005	3,841

D' après le tableau 5.22, le modèle Exponentiel Gamma avec 2 covariés donne la meilleure vraisemblance et ressort comme le meilleur modèle. En effet, selon le tableau 5.22,

- le modèle EG est meilleur que le modèle exponentiel (2459>3,841)
- le modèle EG + 1 covarié (relance par e-mail) est meilleur que le modèle EG (310>3,841)
- le modèle EG + 2 covariés (relance par e-mail/ jour du jeu) est meilleur que le modèle EG+1 covarié (1005>3,841), avec un apport important par rapport au modèle EG+1 covarié.

Le test du ratio de vraisemblance du tableau 5.22 nous permet donc de conclure que le meilleur modèle est le modèle EG+2covariés.

Nous avions également testé un modèle EG+1 covarié (numéro du jour de la semaine) qui donne une vraiseblance de –26191. Le modèle EG+2 covariés est meilleur que ce modèle, le test du ratio de vraisemblance entre 2 covariés et 1 covarié donnant dans ce cas
LR= 903 > 3.84.

Les résultats du modèle EG+2covariés sont fournis dans le tableau excel ci-dessous :

Tableau 5.23 : Paramètres du modèle Exponentiel Gamma + 2 covariés (résultats)

r	13,7594
alpha	22,4930
B_typedejour	-0,8449
B_mail	0,4425

Nbre inscrits
18441

LL= -25739,3943

-0,8 0,44249635

Jours	incr_Trl	Nbre cum	P(T<=t)	P(non reponse	LL	E(T(t))	A(t)	exp(BX)	Covariates type jo	mail	REPONSE PAR REPONSE Observed	Expected
1	11097	11097	0,6018	0,6018	-5636,1821	11097,0025	1,55658815	1,5566	0	1	11097	11097,00
2	3444	14541	0,7726	0,1709	-6084,6688	14248,3550	2,55658815	1	0	0	3444	3151,35
3	858	15399	0,8201	0,0474	-2615,6024	15123,0227	2,98619468	0,4296	1	0	858	874,67
4	511	15910	0,8571	0,0370	-1684,9948	15804,9247	3,41580121	0,4296	1	0	511	681,90
5	604	16514	0,9151	0,0581	-1719,2206	16875,4878	4,41580121	1	0	0	604	1070,56
6	863	17377	0,9486	0,0335	-2930,7001	17493,4252	5,41580121	1	0	0	863	617,94
7	401	17778	0,9683	0,0197	-1574,2197	17857,2225	6,41580121	1	0	0	401	363,80
8	230	18008	0,9802	0,0118	-1020,5415	18075,3751	7,41580121	1	0	0	230	218,15
9	145	18153	0,9874	0,0072	-715,0510	18208,4544	8,41580121	1	0	0	145	133,08
10	86	18239	0,9896	0,0022	-526,9926	18248,6803	8,84540774	0,4296	1	0	86	40,23
11	60	18299	0,9914	0,0018	-379,8123	18281,5361	9,27501428	0,4296	1	0	60	32,86
12	83	18382	0,9944	0,0030	-482,1009	18336,8978	10,2750143	1	0	0	83	55,36
13	59	18441	0,9963	0,0019	-369,3076	18372,1624	11,2750143	1	0	0	59	35,26

Nous retiendrons donc pour modéliser le nombre d'inscriptions de joueurs par jour lié au mail de relance un modèle Exponentiel Gamma avec 2 covariés : le numéro du jour et l'envoi du mail de relance (paramétré 0 /1).
Rappelons encore ici la formule du modèle Exponentiel Gamma avec covariés (Equation 5.1 déjà présentée précédemment) :

$$P(T \leq t) = 1 - \left(\frac{\alpha}{\alpha + A(t)} \right)^{r} \qquad \text{Equation 5.1}$$

avec $A(t) = \Sigma^{t}_{j=1} \exp(\beta' x_i)$

Les paramètres de ce modèle pour le nombre de joueurs issus de l'e-mail de relance sont les suivants (ces paramètres envoient à l'équation 5.1) :

Tableau 5.24: paramètres du modèle EG + 2 covariés
Inscription via mail de relance

r	13,7594
alpha	22,4930
B_typedejour	-0,8449
B_mail	0,4425

r et alpha sont les paramètres du modèle Gamma, B_no est le coefficient β de la variable covariée type de jour de la semaine, B_mail est le coefficient β de la variable covariée envoi du mail de relance. Selon l'équation 5.1, cela signifie que d'une part la probabilité d'arrêt est forte les jours où il y a un e-mail de relance, et d'autre part que la probabilité d'arrêt est plus faible les samedis et dimanches que les autres jours de la semaine (en effet, les samedis et dimanches, A(t) est plus petit).

Durée entre 2 inscriptions mail de relance :

Nous avons adopté la même démarche pour le mail de relance que celle adoptée précédemment pour déterminer la loi de la durée entre 2 inscriptions liées au mail de relance, que nous voulons paramétrer dans le réseau de Petri pour décrire la transition inscription mail de relance.
Le mail de relance ne démarrant que le 27 Mai à 12h22, nous avons inclus dans le réseau de Petri une place et une transition (démarrage inscriptions mail relance) pour intégrer ce délai.

Les principales statistiques de la durée entre 2 inscriptions provenant d'un mail de relance sont fournies dans le tableau 25 ci-dessous :

Tableau 5.25: Statistiques durée entre 2 inscriptions mail relance

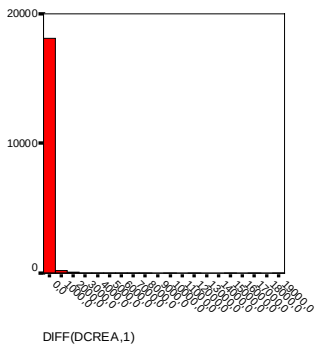
Statistics

DIFF(DCREA,1)

N	Valid	18441
	Missing	1
Mean		57,0502
Median		4,0000
Std. Deviation		370,60410
Variance		137347,4
Minimum		,00
Maximum		19365,00
Percentiles	10	,0000
	20	1,0000
	30	2,0000
	40	3,0000
	50	4,0000
	60	7,0000
	70	14,0000
	80	33,0000
	90	98,0000

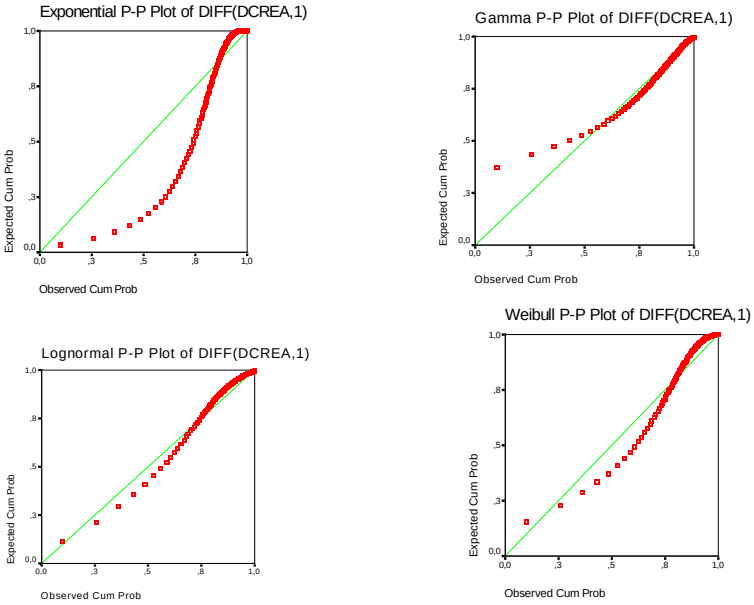
On constate sur le tableau 5.25 que la durée moyenne entre 2 inscriptions venant d'un mail de relance est de 57 secondes, avec des cas extrêmes, le maximum étant à 19365 seconde. L'histogramme de la durée inter inscription venant des mails de relance est présenté dans la figure 5.26.

Figure 5.26 : Histogramme de la durée entre 2 inscriptions mail de relance (en seconde) :



On constate une courbe qui a la même forme décroissante que celles des autres sources de recrutement déjà analysées. Nous avons donc là aussi testé plusieurs lois en analysant tout d'abord les PP Plot. Les résultats sont fournis dans la figure 5.27. La loi exponentielle présente beaucoup d'écarts entre les données observées et le modèle. Nous allons comparer par conséquent les lois Gamma, Weibull et Log-normale.

Figure 5.27 : PP plots durée inter inscription mail de relance



Sur Excel, par la méthode de maximisation de la vraisemblance, nous avons calculé les paramètres de ces différents modèles (les données étant exprimées en secondes). La meilleure vraisemblance est obtenue pour le modèle de Weibull (-25140) versus –25187 pour la loi Gamma et –27448 pour la loi Log-Normale. Même s'il existe peu d' écart de vraisemblance entre les lois Weibull et Gamma, nous choisissons la loi de Weibull dans un souci de cohérence par rapport aux lois déjà sélectionnées pour les autres sources de recrutement. On notera également que la vraisemblance de la loi de Weibull (-25140 avec 2 paramètres) est meilleure que celle du modèle EG + 2 covariés (-25739) qui comporte 4 paramètres.

Comme nous l'avons fait pour le mail d'annonce, et dans un souci de bien représenter les inscriptions par jour issues du mail de relance, nous avons fait varier le paramètre béta de la loi de Weibull dans le temps. Les résultats du modèle sont les suivants :

			alpha Petri		**Pétri**		
			Pétri 10j	1,376905	alpha 2		
			Pétri 20j	1,352147	béta 2	0,2100000	
			Pétri 30j	1,327835			
					LOI WEIBULL EXCEL		
					alpha	0,2100000	
béta excel 10 jrs	0,218048468				lambda0	0,2000000	
béta excel 20 jrs	0,237725671				gamma	0,0000001	
béta excel 30 jrs	0,259178592				LL=	-7286,1052775	
Durée moyenne depuis le début du jeu	Durée Inter inscription	Nbre cas		% réel	P(T=t)		
6,112	200+	9		0,000635324	0,0151989	-37,6787747	
4,6059	190-200	3		0,000211775	0,0002132	-25,3596496	
4,3398	180-190	1		7,05916E-05	0,0007772	-7,1598535	
4,4867	170-180	2		0,000141183	0,0007298	-14,4455140	
4,0288	160-170	2		0,000141183	0,0008655	-14,1042974	
4,0841	150-160	2		0,000141183	0,0008964	-14,0342395	
4,4948	140-150	2		0,000141183	0,0011725	-13,4971963	
3,8582	130-140	2		0,000141183	0,0012018	-13,4478434	
4,4074	120-130	1		7,05916E-05	0,0014166	-6,5594773	
3,6104	110-120	9		0,000635324	0,0014900	-58,5809077	
3,7195	100-110	9		0,000635324	0,0018709	-56,5322207	
3,7808	90-100	8		0,000564732	0,0021147	-49,2707610	
3,6782	80-90	16		0,001129465	0,0025100	-95,7996024	
3,532	70-80	18		0,001270648	0,0030381	-104,3374449	
3,5525	60-70	21		0,001482423	0,0038091	-116,9775345	
3,043	50-60	30		0,002117747	0,0048175	-160,0650900	
2,7354	40-50	62		0,004376677	0,0065106	-312,1282008	
2,3459	30-40	81		0,005717916	0,0092986	-378,9090407	
1,878	20-30	223		0,015741917	0,0150320	-936,0586367	
1,3502	10-20	1001		0,070662149	0,0307953	-3483,8719715	
0,3859	0-10	12664		0,893971481	0,8962412	-1387,2870208	
	total	14166		1			

Ce modèle améliore significativement le niveau de vraisemblance (-29239 versus –29352) avec un seul paramètre supplémentaire et le bon niveau de R^2 (0.98) montre que ce modèle représente bien l'arrivée des joueurs par jour liée au mail d'annonce. C'est par conséquent ce modèle Weibull de paramètres :

$$\alpha \text{ Petri} = (5\ e^{-0,0000001\,t})^{0,21} = 1,40\ e^{-0,000000021\,t}$$

$$\beta \text{ Petri} = 0,21$$

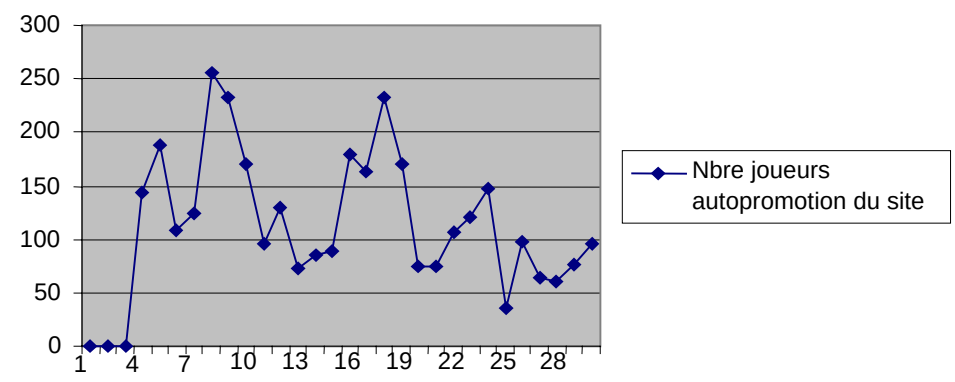
(t étant exprimé en secondes)

que nous retiendrons pour modéliser la durée entre deux inscriptions liées au mail d'annonce. Nous allons appliquer la même démarche pour les inscriptions liées à l'autopromotion sur le site Anyway.

5.4.1.4 Modélisation du nombre d'inscriptions venant de l'auto promotion du site

L'une des provenances de joueurs est l'auto promotion effectuée sur le site Anyway, via une bannière d'annonce du jeu invitant les visiteurs du site Anyway à participer au jeu.
Le nombre d'inscriptions lié à ce stimulus marketing varie d'un jour à l'autre, comme le montre la figure 5.28 ci-dessous. Rappelons cependant que cette source d'inscriptions ne représente que 1% des joueurs.

Figure 5.28: Evolution du nombre de joueurs issus de l'autopromotion du site



Pour modéliser ces inscriptions, tenant compte du fait que ce cas ne représente que peu d'observations relativement aux autres sources d'inscriptions, nous avons étudié un modèle linéaire avec des variations aléatoires autour de la moyenne. Nous avons modélisé ces inscriptions à partir du jour de début de l'auto promotion (4ème jour du jeu), donc sur 27 jours de jeu.
Les résultats de la régression linéaire avec comme variable dépendante le nombre d'inscriptions via autopromotion du site et comme variable indépendante le numéro du jour (de 4 à 30) sont fournis ci-dessous :

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	177,484	19,999		8,875	,000
	jour	-3,696	1,248	-,510	-2,961	,007

a. Dependent Variable: nbre inscrits autopromotion site

En acceptant un risque α de 10%, on peut conclure que le nombre d'inscriptions est lié au numéro du jour du jeu.
Le modèle utilisé est donc :

Nbre d'inscriptions autopromotion = -3.696 *(numéro du jour) + 177,484

Ce modèle donne un R^2 de 0.260, ce qui est faible. Cependant, nous ne cherchons dans le cas de l'autopromotion du site qu'à avoir une tendance générale des inscriptions issues de cette source. Il est clair que ce modèle présente donc quelques limites puisqu'il ne représente pas correctement les données par jour (il représente la tendance générale). Cependant, il est difficile sur de l'autopromotion de site de trouver d'autres variables explicatives et disponibles permettant de modéliser ce phénomène (on pourrait évoquer le nombre de visiteurs du site, mais cette donnée appartenant à Anyway ne nous a pas été fournie). D'autre part, les inscriptions liées à l'autopromotion du site ne représentent que 1% des joueurs, ce qui minimise l'impact du modèle sélectionné ici.

Durée entre 2 inscriptions autopromotion du site :

Nous cherchons ici à modéliser la loi de la durée entre 2 inscriptions liées à l'autopromotion sur le site Anyway. La première inscription liée à cette source de recrutement s'effectue le 13 mai ($4^{ème}$ jour du jeu) à 14h55 donc nous avons ajouté dans le réseau de Petri, comme nous l'avons fait pour les sources de recrutement précédentes, une transition (démarrage inscription autopromotion) et une place pour tenir compte de ce délai.

Tableau 5.27 : Statistiques durée inter inscription autopromotion du site :

Statistics

DIFF(DCREA,1)

N	Valid	3395
	Missing	1
Mean		655,7418
Median		282,0000
Std. Deviation		1914,500
Variance		3665312
Minimum		,00
Maximum		74571,00
Percentiles	10	39,0000
	20	80,0000
	30	131,0000
	40	197,0000
	50	282,0000
	60	390,0000
	70	548,0000
	80	776,0000
	90	1301,0000

Le tableau 5.27 montre une moyenne de durée inter inscription de 655 secondes, soit un délai bien supérieur à celui constaté sur les sources de recrutement précédentes. On a déjà globalement constaté que cette source de recrutement représente une faible part des recrutements pour le jeu Anyway. L'histogramme des données est représenté dans la figure 5.28 ci-dessous :

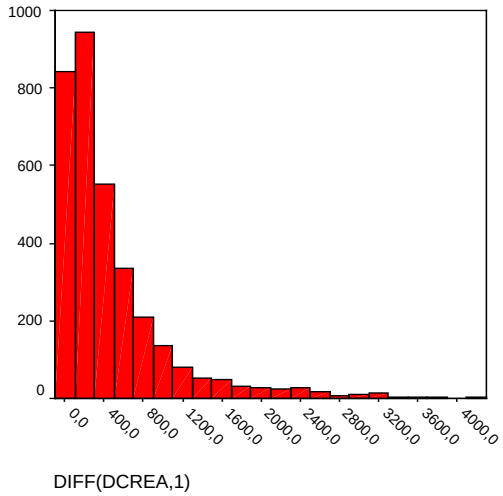
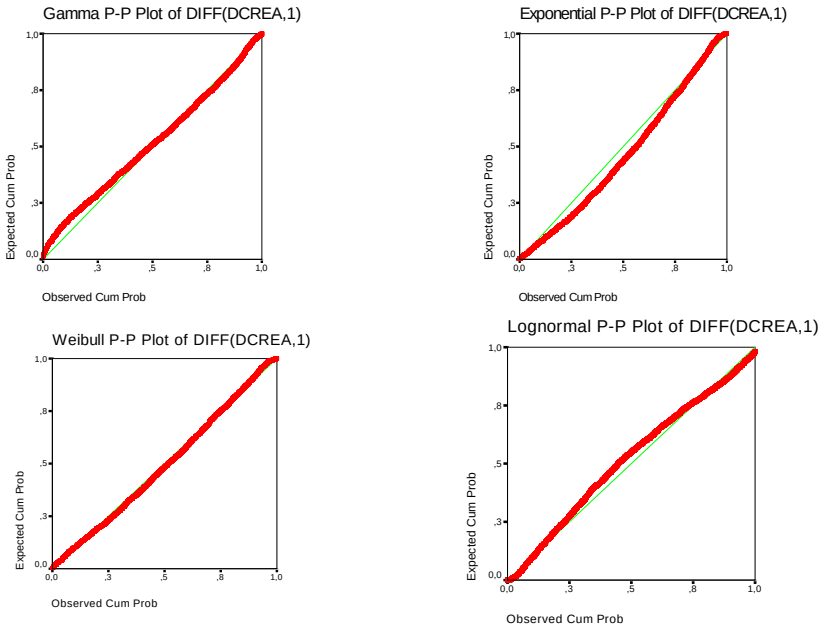


Figure 5.28 : histogramme des durées inter inscription autopromotion

Comme pour les sources de recrutement précédentes, nous avons donc testé plusieurs lois en regardant tout d'abord les PP Plots.
Les lois les plus proches sont représentées sur la figure 5.29. On constate encore une fois que les lois les plus représentatives des données sont les lois Gamma, Weibull et Log-Normale.

Figure 5.29 : PP Plots durée inter inscription autopromotion



Après estimation des paramètres de ces lois sur excel (par la maximisation de la vraisemblance), on obtient une vraisemblance de –6406 pour la loi Gamma contre –6413 pour la loi Log-Normale et –6405 pour la loi de Weibull. Nous avons donc sélectionné pour la durée inter inscription venant de l'autopromotion sur le site une loi de Weibull dont les paramètres sont les suivants :

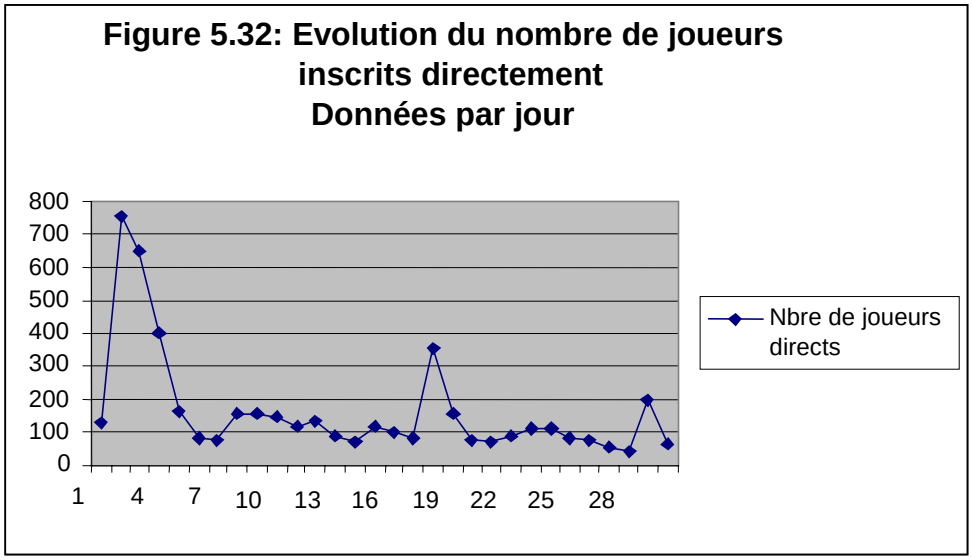
Tableau 5.28: Paramètres Pétri de la loi Weibull autopromotion

alpha 4	0,0084
bêta 4	0,7602

5.4.1.5 Modélisation du nombre d'inscriptions d'origine non identifiée

Pour le jeu « Le ciel est à tout le monde », 2,1% des participants se sont inscrits sans préciser par quel moyen ils sont venus au jeu.

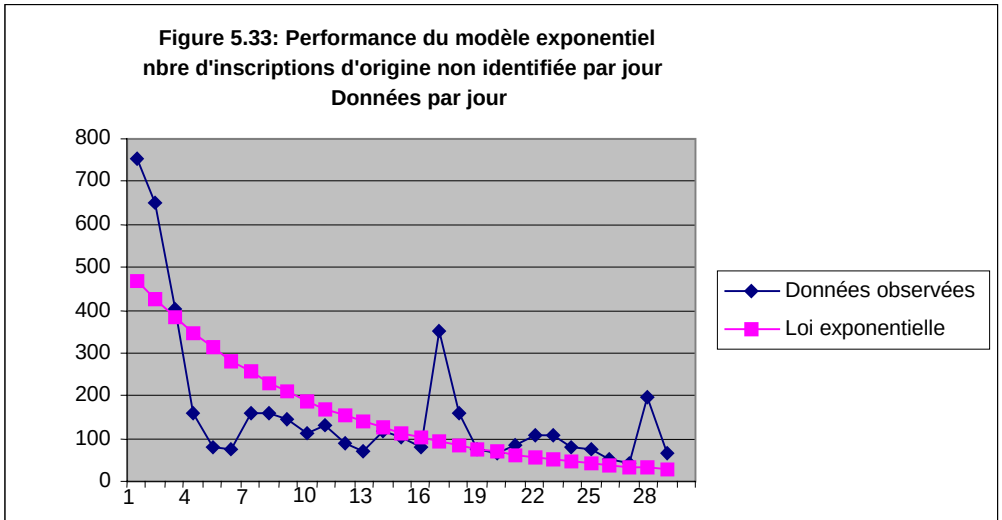
L'évolution de ces inscriptions pour chaque jour du jeu est représentée sur la figure 5.32 ci-dessous.



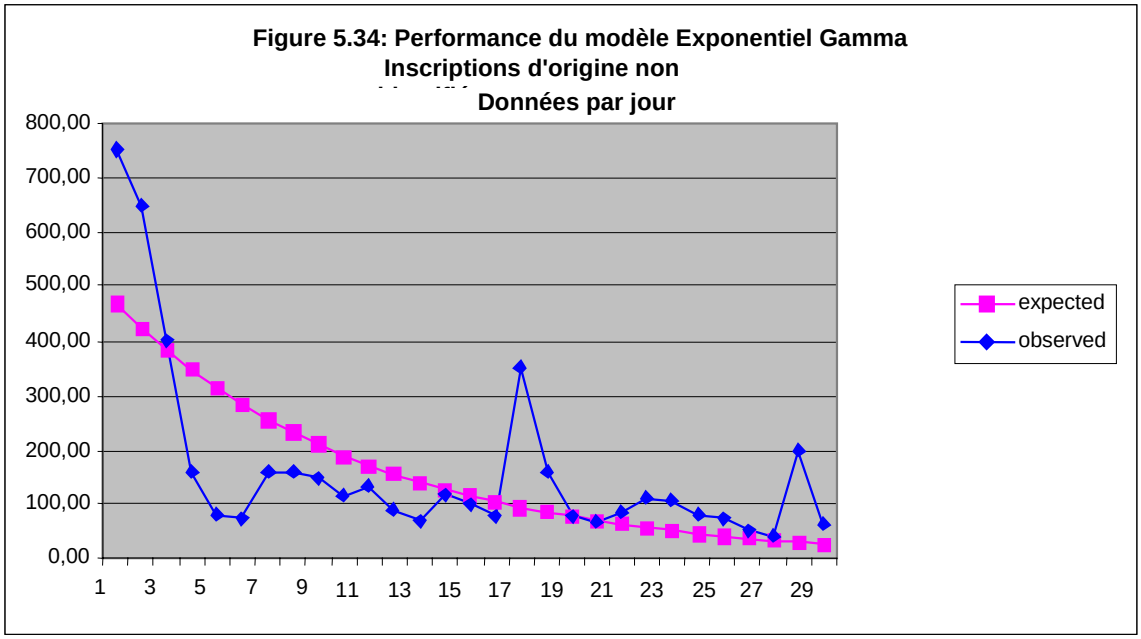
On constate que cette courbe présente des pics en début de jeu (probablement liés au mail d'annonce et à l'affiliation, les internautes ayant par bouche à oreille oral parlé du jeu à d'autres internautes) puis lors du mail de relance.

Nous allons modéliser le nombre d'inscriptions de ce type selon la même démarche que celle que nous avons utilisée dan les modèles précédents : nous avons donc testé là aussi un modèle exponentiel, exponentiel gamma et EG avec covariés, en travaillant uniquement sur les données à partir du 2ème jour du jour, comme nous l'avons fait pour l'affiliation ou le mail d'annonce.

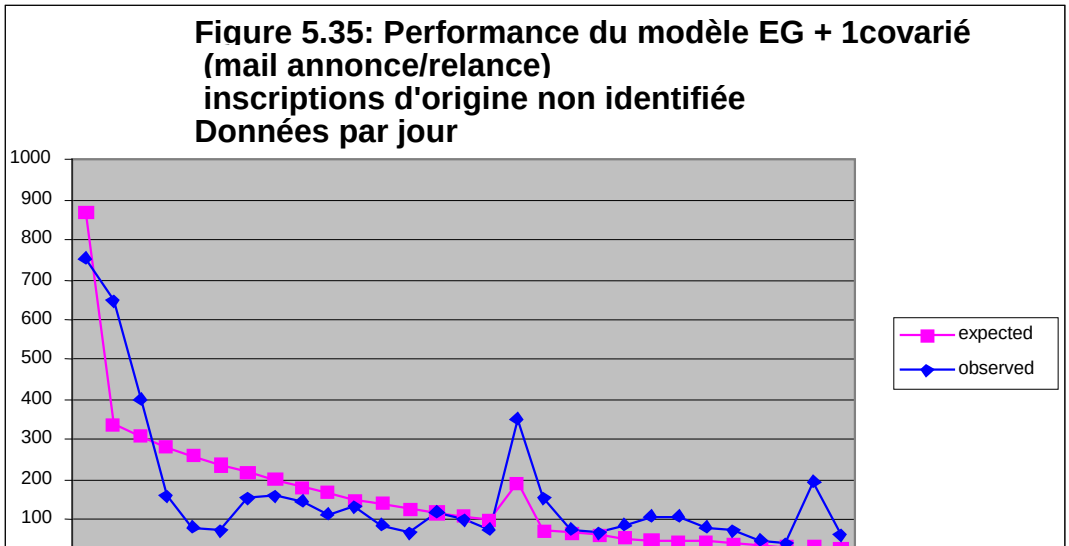
La performance du modèle exponentiel est la suivante :



Lorsqu'on introduit dans le modèle l'hétérogénéité entre les individus et qu'on teste un modèle Exponentiel Gamma, les résultats sont les suivants :

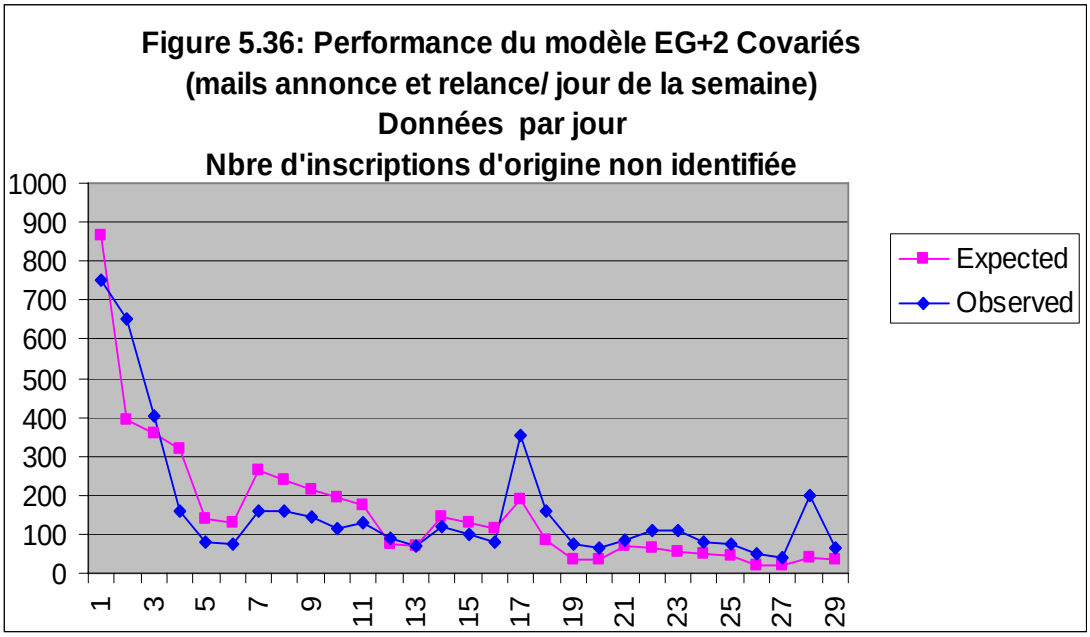


Comme le montre la figure 5.34, le modèle Exponentiel Gamma n'intègre pas les pics de début de jeu et le pic d'inscriptions probablement dû au mail de relance. Afin d'améliorer le modèle, nous avons donc testé un modèle Exponentiel Gamma avec un covarié. La variable covariée choisie est l'envoi d'e-mail (annonce ou relance), introduite de manière binaire (1 envoi/ 0 non envoi). Cette variable vaut donc 1 pour le 1er jour et le 18ème jour du jeu, 0 pour les autres jours. Les résultats de ce modèle sont présentés dans la figure 5.35 :



On constate une amélioration des données estimées versus les données observées mais avec encore des écarts sur certains jours du jeu. Nous avons donc introduit une deuxième variable covariée qui est le type de jour de la semaine, comme nous l'avons déjà fait sur les modèles précédents, en faisant l'hypothèse que le comportement des internautes n'est pas le même en semaine et le week-end (cette variable est paramétrée 0 du lundi au vendredi et 1 les samedis et dimanches).

Les résultats sont les suivants :



Afin de déterminer le meilleur de ces modèles, nus avons utilisé le test du ratio de vraisemblance, présent dans le tableau 5.29 ci-dessous.

Tableau 5.29: Comparaison des modèles

			Comparaison du modèle vs modèle ligne précédente	
	LL	Nbre de paramètres	LR	Khi 2 (0,05;1)
Exponentiel	-15620,57	1		
EG	-15620,57	2	0	3,841
EG+ 1 covarié	-15540,56	3	160	3,841
EG+ 2 covariés	-15226,89	4	627	3,841

La conclusion de ce tableau est que le meilleur modèle est l'Exponentiel Gamma avec 2 covariés. En effet, on constate que :

- Le modèle EG n'apporte rien par rapport au modèle exponentiel (0<3,3841)
- Le modèle EG + 1 covarié est meilleur que le modèle EG (160 >3,3841)
- Le modèle EG + 2 covariés est meilleur que le modèle EG+1covarié (627>3,3841)

Les paramètres du modèle EG + 2 Covariés sont les suivants :

Tableau 5.30: Paramètres du modèle EG+2covariés inscription source non identifiée

r	89950,6395
alpha	867513,7111
B_mail	0,6354
B_jour	-0,7677

r et alpha sont les paramètres du modèle Gamma, B_mail est le coefficient de la variable covariée e-mail (annonce et relance), B_jour est le coefficient de la variable type de jour de la semaine. Le modèle se réfère à l'équation 5.1.

Nous retenons ce modèle même s'il comporte certains écarts par rapport aux données observées car les inscriptions d'origine non identifiée ne représentent que 2% des inscriptions totales. L'impact de l'imperfection du modèle sera donc faible sur l'ensemble des données observées d'inscriptions de joueurs.

Durée inter inscription source non identifiée :

Nous cherchons dans cette partie à modéliser la loi de la durée entre 2 inscriptions liées à une source non identifiée, afin de paramétrer cette loi dans le réseau de Petri pour la transition inscription liée à une source non identifiée.
La première inscription liée à une source non identifiée s'est effectuée le premier jour du jour à 13h47. Nous avons donc ajouté dans le réseau de Petri, comme nous l'avons fait pour les sources de recrutement précédentes, une transition (démarrage inscription source non identifiée) et une place permettant d'intégrer ce délai.

Tableau 5.32 : Statistiques durée inter inscription source non identifiée

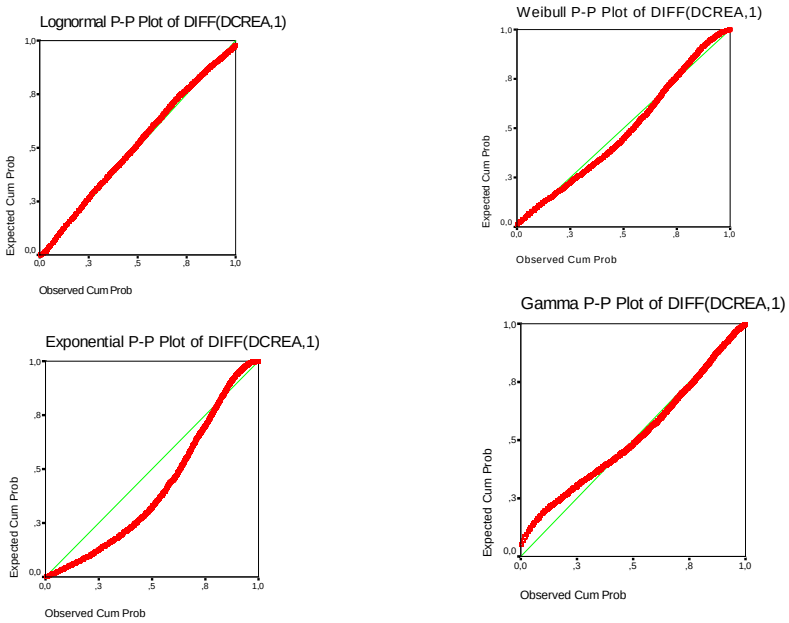
Statistics

DIFF(DCREA,1)

N	Valid	4873
	Missing	1
Mean		509,0432
Median		148,0000
Std. Deviation		1334,471
Variance		1780813
Minimum		,00
Maximum		22289,00
Percentiles	10	17,0000
	20	36,0000
	30	63,0000
	40	97,0000
	50	148,0000
	60	228,0000
	70	379,0000
	80	610,0000
	90	1146,0000

Il y a en moyenne une inscription venant d'une source non identifiée toutes les 509 secondes. Là aussi, cette moyenne est influencée par des cas extrêmes : le maximum est à 22289 secondes.
Les résultats des PP plots des lois ensuite testées sont présentés sur la figure 5.37.
On retient encore une fois les trois lois Gamma, Weibull et Log-Normale.

Figure 5.37 : PP plots durée inter inscription source non identifiée



Pour sélectionner la meilleure de ces 3 lois, nous avons estimé les paramètres sous excel. La vraisemblance du modèle de Weibull est de –11309 versus –11304 pour la loi Log-Normale et –11704 pour la loi Gamma.

Pour des raisons d'homogénéité avec les résultats des sources précédentes et en raisons des très faibles écarts de vraisemblance entre les lois Log-Normale et Weibull, nous retenons donc pour modéliser la durée entre 2 inscriptions de source non identifiée une loi de Weibull (Equation 2) de paramètres :

Tableau 5.33: Paramètres Pétri loi de Weibull source non identifiée

alpha 5	0,0125
béta 5	0,7225

Nous avons par conséquent modélisé dans cette partie 5.4.1 le nombre d'inscriptions de joueurs selon les différents stimuli marketing.

Nous allons dans la partie suivante s'intéresser aux réponses des joueurs pour chaque question du jeu, afin de déterminer la probabilité des transitions Q1, Q2, etc… du réseau de Petri de la figure 5.2.

5.4.2 Modèle des non-réponses aux questions :

L'objectif de cette partie est de déterminer les lois de chaque transition *Q1, Q2, Q3, ..., Q41* du réseau de Petri, ou encore le nombre de réponses et de non-réponses pour chaque question. Nous avons choisi de modéliser le nombre de non-réponses par question (on en tire ensuite simplement le nombre de réponses).

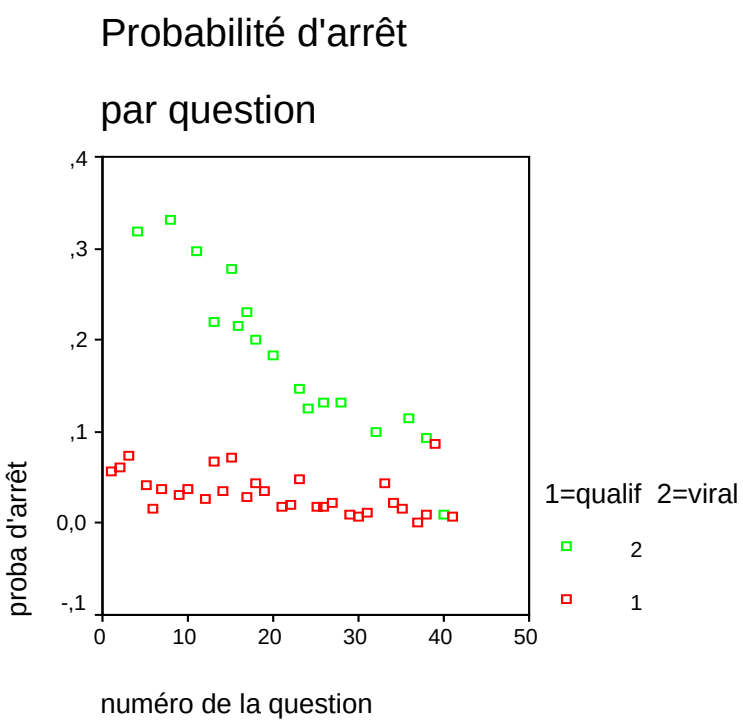
Les données brutes observées pour chaque question sur la probabilité qu'un joueur s'arrête, en fonction du numéro de la question et du type de question (qualification ou virale), sont fournies ci-dessous et représentées dans la figure 5.38.

Données brutes des réponses aux questions

Nbre envois	Nbre réponses	Q	Non réponses	Type de question	Proba d'arrêt
227879	214768	1	13111	1	0,0575
214768	201622	2	13146	1	0,0612
201622	186893	3	14729	1	0,0731
186893	127375	4	59518	2	0,3185
127375	122159	5	5216	1	0,0409
122159	120263	6	1896	1	0,0155
120263	115717	7	4546	1	0,0378
115717	77368	8	38349	2	0,3314
77368	74937	9	2431	1	0,0314
74937	72127	10	2810	1	0,0375
72127	50667	11	21460	2	0,2975
50667	49362	12	1305	1	0,0258
24551	22872	13Q	1679	1	0,0684
24811	19354	13V	5457	2	0,2199
42226	40728	14	1498	1	0,0355
25501	23668	15Q	1833	1	0,0719
15227	10999	15V	4228	2	0,2777
34667	27179	16	7488	2	0,2160
18325	17794	17Q	531	1	0,0290
8854	6806	17V	2048	2	0,2313
7273	6959	18Q	314	1	0,0432
17327	13865	18V	3462	2	0,1998
20824	20105	19	719	1	0,0345
20105	16413	20	3692	2	0,1836
16413	16126	21	287	1	0,0175
16126	15797	22	329	1	0,0204
4113	3917	23Q	196	1	0,0477
11624	9914	23V	1710	2	0,1471
13831	12093	24	1738	2	0,1257
12093	11879	25	214	1	0,0177
10130	9953	26Q	177	1	0,0175
1749	1520	26V	229	2	0,1309
11473	11226	27	247	1	0,0215
11226	9751	28	1475	2	0,1314
9751	9652	29	99	1	0,0102
9652	9572	30	80	1	0,0083
9572	9471	31	101	1	0,0106
9471	8520	32	951	2	0,1004
8520	8152	33	368	1	0,0432
8152	7975	34	177	1	0,0217
7975	7855	35	120	1	0,0150
7855	6947	36	908	2	0,1156
6947	6934	37	13	1	0,0019
5733	5682	38Q	51	1	0,0089
1201	1090	38V	111	2	0,0924
6772	6179	39	593	1	0,0876
6179	6121	40	58	2	0,0094
6121	6074	41	47	1	0,0077

1=qualif
2=viral

Figure 5.38 : Données observées sur le nombre de non répondants par question



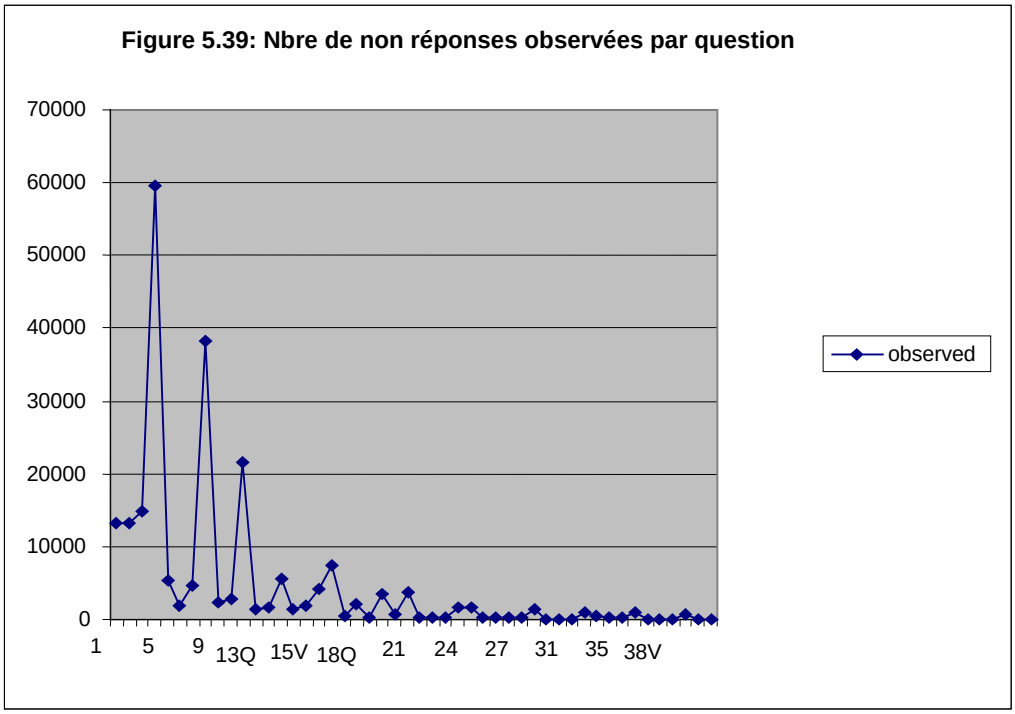
On constate deux phénomènes dans ce graphique. Le premier est qu'il existe une forte différence dans le comportement de réponse des joueurs selon qu'ils reçoivent une question de qualification (dans ce cas, selon les questions, entre 0 et 10% des joueurs ne continuent pas le jeu) ou une question virale (entre 8 et 33% des joueurs s'arrêtent). Il semble donc intéressant d'analyser séparément les réponses aux questions de qualification et aux questions virales. Le deuxième phénomène à noter est que la probabilité d'arrêt décroît avec le nombre de questions. On aurait pu s'attendre logiquement plutôt à l'inverse : plus le joueur avance dans le questionnaire, moins il a d'adresses e-mail connues à fournir ou plus il se « lasse » des questions, donc la probabilité d'arrêt devrait augmenter. Une autre explication peut venir de l'hétérogénéité entre les individus : ceux qui abandonnent dès les premières questions n'étaient pas très « joueurs » ou ont peu d'adresses e-mail à fournir. En revanche, ceux qui restent jusqu'aux dernières questions sont les plus impliqués par le jeu, ou ont plus d'adresses e-mail à donner, ce noyau ayant par conséquent une probabilité d'arrêt faible. Cette hypothèse suppose qu'il faudra probablement intégrer dans notre modèle l'hétérogénéité des individus.

Pour modéliser le nombre de non-réponses par question, nous allons développer deux approches différentes. La première propose un modèle continu fondé sur l'hypothèse selon laquelle les actions séquentielles du joueur (il répond à la question Q1, puis Q2, etc… jusqu'à la question Q_n à laquelle il s'arrête) constituent un phénomène continu. La deuxième approche repose sur un modèle discret selon lequel la probabilité de non-réponse à chaque question correspond à un processus binaire (réponse/non-réponse).

5.4.2.1 Modèle continu de non-réponse aux questions :

Nous cherchons à modéliser la probabilité d'arrêt pour chaque question. La figure 5.39 représente le nombre observé de non-réponses par question. On retrouve ici les fortes différences entre questions virales (ex : Q4,Q8,etc…) pour lesquelles on constate des pics de non-réponses, et les questions de qualification pour lesquelles la probabilité d'arrêt est moins forte.

Figure 5.39: Nbre de non réponses observées par question



Modèle Exponentiel-Gamma avec covariés :

Etant donné que, selon l'analyse de la figure 5.38 précédemment décrite, nous voulons intégrer dans notre modèle le type de question et l'hétérogénéité des individus (qui a priori influent sur la probabilité de non réponse), nous avons donc dans un premier temps modélisé la probabilité de non réponses par question par un modèle Exponentiel Gamma avec covariés (nous avons testé au préalable un modèle exponentiel et un modèle Exponentiel Gamma qui n'ont pas donné de bons résultats). Selon ce modèle, la probabilité de non réponse à la question x suit une loi exponentielle de paramètre lambda, ce paramètre étant distribué sur l'ensemble des individus selon une loi Gamma.

Nous avons également introduit des variables covariées afin d'intégrer dans notre modèle des différences selon les questions posées. En effet, intuitivement, on s'attend à ce que la probabilité de non réponse d'un joueur à une question donnée soit fonction des caractéristiques du questionnaire, comme par exemple le type de question proposée (virale ou qualification), le choix de l'ordre des questions ou encore la complexité de la réponse (nbre de choix possibles de réponses). Ces caractéristiques seront introduites dans le modèle comme covariées.

Rappel sur l'intégration de covariés (nous avons déjà appliqué cette démarche précédemment pour la modélisation des incriptions au jeu):

Dans le cas du modèle Exponentiel-Gamma avec covariés, on incorpore les variables covariées de la manière suivante:

- Soit x(t) le vecteur des variables covariées x_i au temps t et β les effets de ces covariés

- On fait l'hypothèse que les covariés ont un effet multiplicatif sur la base-line de la fonction de hasard suivante, définie pour la distribution exponentielle :

  $$h(t/\lambda, x(t), \beta) = \lambda \exp(\beta' x(t))$$

  Cette fonction de taux de hasard représente le taux instantané de « défaillance » en t conditionnel à la « survie » en t.

$$h(t)=\lim_{\Delta t\to 0} P(t<T\leq t+\Delta t \mid T>t)/\Delta t=\frac{f(t)}{1-F(t)}$$

On peut dériver la distribution correspondante (avec les covariés) en reprenant la relation fondamentale entre la fonction de hasard et la cdf d'une distribution :

$$F(t)=1-\exp\left(\int_0^t h(u)du\right)$$

Cette fonction F(t) est nommée régression proportionnelle de hasard

- La distribution exponentielle a un taux de hasard constant λ ($F(t)=1-e^{-\lambda t}$)

Donc en incluant les covariés, on obtient :

$$F(t) = 1 - e^{-\lambda A(t)}$$

Avec $A(t) = \Sigma^{t}_{j=1} \exp(\beta' x_i)$

- Dans le modèle exponentiel gamma, le paramètre lambda est distribué sur l'ensemble des joueurs selon une distribution gamma :

$$g(\lambda)=\frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}$$

où r est le paramètre de forme du modèle et α le paramètre d'échelle

- La distribution cumulée du modèle Exponentiel-Gamma avec covariés est:

$$P(T\le t)=\int_0^\infty (1-\exp(-\lambda A(t)))\frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}d\lambda$$

D'où $$P(T\le t)=1-\left(\frac{\alpha}{\alpha+A(t)}\right)^r$$ Equation 5.1

avec $A(t) = \Sigma^{t}_{j=1} \exp(\beta' x_i)$

Plusieurs options de modèles ont été testées :

1. Modèle EG + 1 covarié (type de question)
2. Modèle EG + 2 covariés (type de question + Numéro de question)
3. Modèle EG + 3 covariés (type de question + Numéro de question+ Nbre de questions virales déjà répondues)
4. Modèle EG + 4 covariés (type de question + Numéro de question+ Nbre de questions virales déjà répondues + complexité de la question)

Les formules et les résultats de ces 4 modèles sont fournis dans les tableaux ci-dessous. Les paramètres sont estimés par la méthode du maximum de vraisemblance sous Excel (Hardie et Fader, 2002). Les performances des modèles sont visualisées dans chaque cas sur les données non cumulées par question.

Modèle EG + 1 covarié (Type de question) :

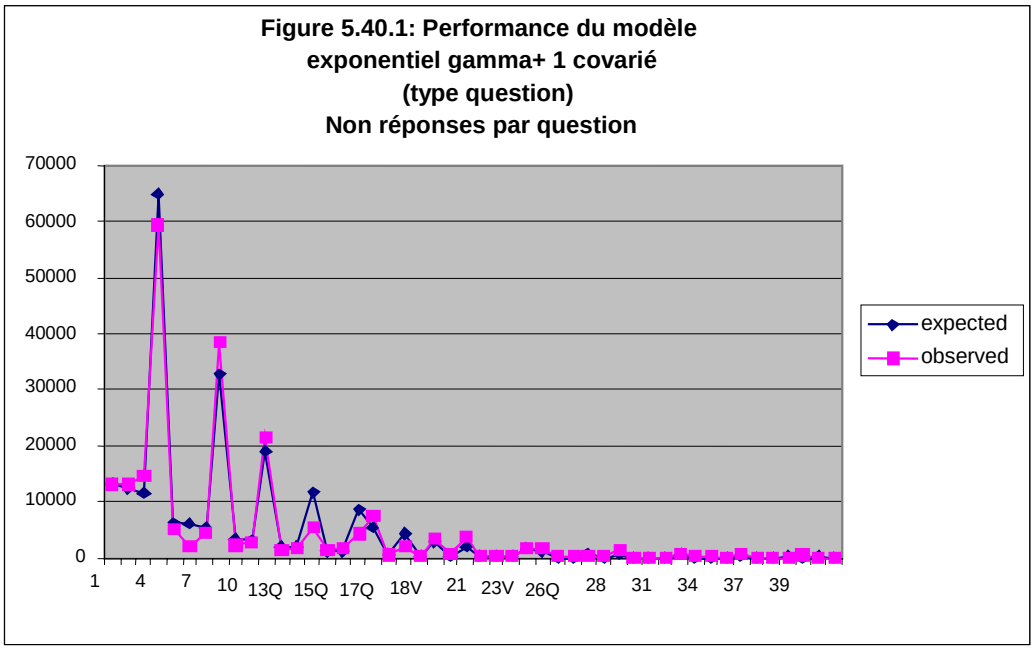
La première variable covariée qu'il nous semble importante d'intégrer dans le modèle est effectivement le type de question (virale ou de qualification) qui, comme nous l'avons constaté dans la figure 5.38, semble influer sur la probabilité de réponse aux questions.

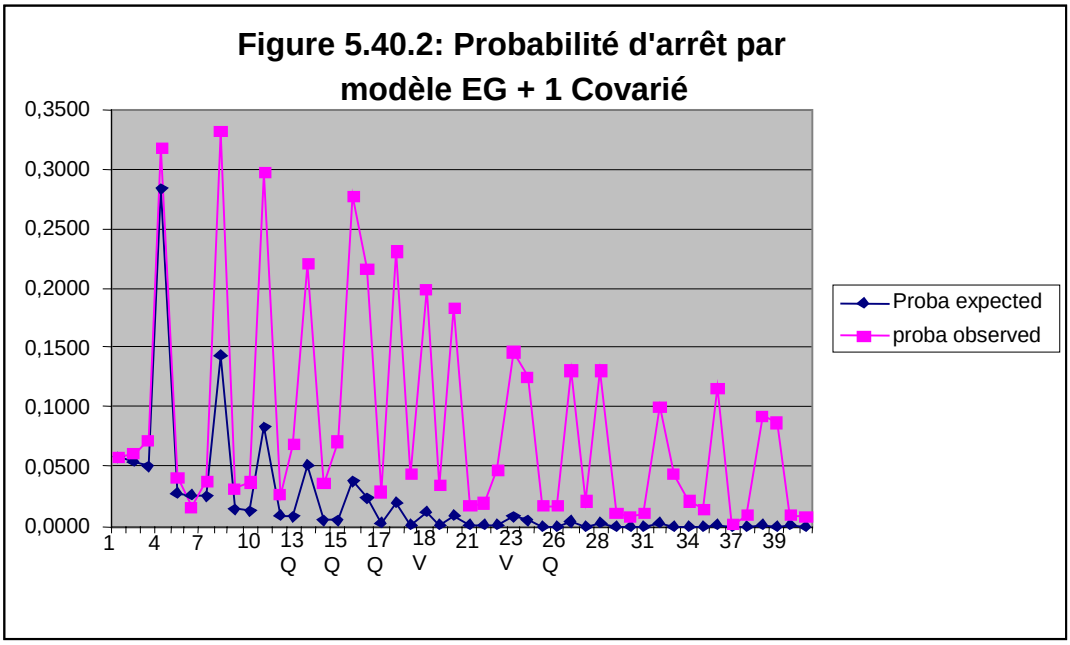
Tableau 5.34 : Modèle EG + 1 covarié (type de question) : Formules

nbre tc	227879				r	3,828878686816						
de					alpha	475,6870788249						
joueur:					B_no	2,021819830858						
					LL=	=SOMME(F7:F5		=TRAN				
Questi	Nbre envois	Nbre non réponses	P(T<=t)	(Non repons	LL	E(T(t))	A(t)	exp(BX)	Covar. type	PAR Observed	QUESTION Expected	
1	227879	13111	=1-(F$2/(F$2+J7))^F$1	=D7	=C7*LN(E7)	=B$1*D7	=K7	=EXP(SOMMEPROD(I$4;L7))	1	=C7	=B$1*E7	
2	214768	13146	=1-(F$2/(F$2+J8))^F$1	=D8-D7	=C8*LN(E8)	=B$1*D8	=J7+K8	=EXP(SOMMEPROD(I$4;L8))	1	=C8	=B$1*E8	
3	201622	14729	=1-(F$2/(F$2+J9))^F$1	=D9-D8	=C9*LN(E9)	=B$1*D9	=J8+K9	=EXP(SOMMEPROD(I$4;L9))	1	=C9	=B$1*E9	
4	186893	59518	=1-(F$2/(F$2+J10))^F$1	=D10-D9	=C10*LN(E10)	=B$1*D10	=J9+K10	=EXP(SOMMEPROD(I$4;L10))	2	=C10	=B$1*E10	
5	127375	5216	=1-(F$2/(F$2+J11))^F$1	=D11-D10	=C11*LN(E11)	=B$1*D11	=J10+K11	=EXP(SOMMEPROD(I$4;L11))	1	=C11	=B$1*E11	
6	122159	1896	=1-(F$2/(F$2+J12))^F$1	=D12-D11	=C12*LN(E12)	=B$1*D12	=J11+K12	=EXP(SOMMEPROD(I$4;L12))	1	=C12	=B$1*E12	
7	120263	4546	=1-(F$2/(F$2+J13))^F$1	=D13-D12	=C13*LN(E13)	=B$1*D13	=J12+K13	=EXP(SOMMEPROD(I$4;L13))	1	=C13	=B$1*E13	
8	115717	38349	=1-(F$2/(F$2+J14))^F$1	=D14-D13	=C14*LN(E14)	=B$1*D14	=J13+K14	=EXP(SOMMEPROD(I$4;L14))	2	=C14	=B$1*E14	
9	77368	2431	=1-(F$2/(F$2+J15))^F$1	=D15-D14	=C15*LN(E15)	=B$1*D15	=J14+K15	=EXP(SOMMEPROD(I$4;L15))	1	=C15	=B$1*E15	
10	74937	2810	=1-(F$2/(F$2+J16))^F$1	=D16-D15	=C16*LN(E16)	=B$1*D16	=J15+K16	=EXP(SOMMEPROD(I$4;L16))	1	=C16	=B$1*E16	
11	72127	21460	=1-(F$2/(F$2+J17))^F$1	=D17-D16	=C17*LN(E17)	=B$1*D17	=J16+K17	=EXP(SOMMEPROD(I$4;L17))	2	=C17	=B$1*E17	
12	50667	1305	=1-(F$2/(F$2+J18))^F$1	=D18-D17	=C18*LN(E18)	=B$1*D18	=J17+K18	=EXP(SOMMEPROD(I$4;L18))	1	=C18	=B$1*E18	
13Q	24551	1679	=1-(F$2/(F$2+J19))^F$1	=D19-D18	=C19*LN(E19)	=B$1*D19	=J18+K19	=EXP(SOMMEPROD(I$4;L19))	1	=C19	=B$1*E19	
13V	24811	5457	=1-(F$2/(F$2+J20))^F$1	=D20-D19	=C20*LN(E20)	=B$1*D20	=J19+K20	=EXP(SOMMEPROD(I$4;L20))	2	=C20	=B$1*E20	
14	42226	1498	=1-(F$2/(F$2+J21))^F$1	=D21-D20	=C21*LN(E21)	=B$1*D21	=J20+K21	=EXP(SOMMEPROD(I$4;L21))	1	=C21	=B$1*E21	
15Q	25501	1833	=1-(F$2/(F$2+J22))^F$1	=D22-D21	=C22*LN(E22)	=B$1*D22	=J21+K22	=EXP(SOMMEPROD(I$4;L22))	1	=C22	=B$1*E22	
15V	15227	4228	=1-(F$2/(F$2+J23))^F$1	=D23-D21	=C23*LN(E23)	=B$1*D23	=J22+K23	=EXP(SOMMEPROD(I$4;L23))	2	=C23	=B$1*E23	
16	34667	7488	=1-(F$2/(F$2+J24))^F$1	=D24-D23	=C24*LN(E24)	=B$1*D24	=J23+K24	=EXP(SOMMEPROD(I$4;L24))	2	=C24	=B$1*E24	
17Q	18325	531	=1-(F$2/(F$2+J25))^F$1	=D25-D24	=C25*LN(E25)	=B$1*D25	=J24+K25	=EXP(SOMMEPROD(I$4;L25))	1	=C25	=B$1*E25	
17V	8854	2048	=1-(F$2/(F$2+J26))^F$1	=D26-D24	=C26*LN(E26)	=B$1*D26	=J25+K26	=EXP(SOMMEPROD(I$4;L26))	2	=C26	=B$1*E26	
18Q	7273	314	=1-(F$2/(F$2+J27))^F$1	=D27-D26	=C27*LN(E27)	=B$1*D27	=J26+K27	=EXP(SOMMEPROD(I$4;L27))	1	=C27	=B$1*E27	
18V	17327	3462	=1-(F$2/(F$2+J28))^F$1	=D28-D27	=C28*LN(E28)	=B$1*D28	=J27+K28	=EXP(SOMMEPROD(I$4;L28))	2	=C28	=B$1*E28	
19	20824	719	=1-(F$2/(F$2+J29))^F$1	=D29-D28	=C29*LN(E29)	=B$1*D29	=J28+K29	=EXP(SOMMEPROD(I$4;L29))	1	=C29	=B$1*E29	
20	20105	3692	=1-(F$2/(F$2+J30))^F$1	=D30-D29	=C30*LN(E30)	=B$1*D30	=J29+K30	=EXP(SOMMEPROD(I$4;L30))	2	=C30	=B$1*E30	
21	16413	287	=1-(F$2/(F$2+J31))^F$1	=D31-D30	=C31*LN(E31)	=B$1*D31	=J30+K31	=EXP(SOMMEPROD(I$4;L31))	1	=C31	=B$1*E31	
22	16126	329	=1-(F$2/(F$2+J32))^F$1	=D32-D31	=C32*LN(E32)	=B$1*D32	=J31+K32	=EXP(SOMMEPROD(I$4;L32))	1	=C32	=B$1*E32	
23Q	4113	196	=1-(F$2/(F$2+J33))^F$1	=D33-D32	=C33*LN(E33)	=B$1*D33	=J32+K33	=EXP(SOMMEPROD(I$4;L33))	1	=C33	=B$1*E33	
23V	11624	1710	=1-(F$2/(F$2+J34))^F$1	=D34-D32	=C34*LN(E34)	=B$1*D34	=J33+K34	=EXP(SOMMEPROD(I$4;L34))	2	=C34	=B$1*E34	
24	13831	1738	=1-(F$2/(F$2+J35))^F$1	=D35-D34	=C35*LN(E35)	=B$1*D35	=J34+K35	=EXP(SOMMEPROD(I$4;L35))	2	=C35	=B$1*E35	
25	12093	214	=1-(F$2/(F$2+J36))^F$1	=D36-D35	=C36*LN(E36)	=B$1*D36	=J35+K36	=EXP(SOMMEPROD(I$4;L36))	1	=C36	=B$1*E36	
26Q	10130	177	=1-(F$2/(F$2+J37))^F$1	=D37-D36	=C37*LN(E37)	=B$1*D37	=J36+K37	=EXP(SOMMEPROD(I$4;L37))	1	=C37	=B$1*E37	
26V	1749	229	=1-(F$2/(F$2+J38))^F$1	=D38-D36	=C38*LN(E38)	=B$1*D38	=J37+K38	=EXP(SOMMEPROD(I$4;L38))	2	=C38	=B$1*E38	
27	11473	247	=1-(F$2/(F$2+J39))^F$1	=D39-D38	=C39*LN(E39)	=B$1*D39	=J38+K39	=EXP(SOMMEPROD(I$4;L39))	1	=C39	=B$1*E39	
28	11226	1475	=1-(F$2/(F$2+J40))^F$1	=D40-D39	=C40*LN(E40)	=B$1*D40	=J39+K40	=EXP(SOMMEPROD(I$4;L40))	2	=C40	=B$1*E40	
29	9751	99	=1-(F$2/(F$2+J41))^F$1	=D41-D40	=C41*LN(E41)	=B$1*D41	=J40+K41	=EXP(SOMMEPROD(I$4;L41))	1	=C41	=B$1*E41	
30	9652	80	=1-(F$2/(F$2+J42))^F$1	=D42-D41	=C42*LN(E42)	=B$1*D42	=J41+K42	=EXP(SOMMEPROD(I$4;L42))	1	=C42	=B$1*E42	
31	9572	101	=1-(F$2/(F$2+J43))^F$1	=D43-D42	=C43*LN(E43)	=B$1*D43	=J42+K43	=EXP(SOMMEPROD(I$4;L43))	1	=C43	=B$1*E43	
32	9471	951	=1-(F$2/(F$2+J44))^F$1	=D44-D43	=C44*LN(E44)	=B$1*D44	=J43+K44	=EXP(SOMMEPROD(I$4;L44))	2	=C44	=B$1*E44	
33	8520	368	=1-(F$2/(F$2+J45))^F$1	=D45-D44	=C45*LN(E45)	=B$1*D45	=J44+K45	=EXP(SOMMEPROD(I$4;L45))	1	=C45	=B$1*E45	
34	8152	177	=1-(F$2/(F$2+J46))^F$1	=D46-D45	=C46*LN(E46)	=B$1*D46	=J45+K46	=EXP(SOMMEPROD(I$4;L46))	1	=C46	=B$1*E46	
35	7975	120	=1-(F$2/(F$2+J47))^F$1	=D47-D46	=C47*LN(E47)	=B$1*D47	=J46+K47	=EXP(SOMMEPROD(I$4;L47))	1	=C47	=B$1*E47	
36	7855	908	=1-(F$2/(F$2+J48))^F$1	=D48-D47	=C48*LN(E48)	=B$1*D48	=J47+K48	=EXP(SOMMEPROD(I$4;L48))	2	=C48	=B$1*E48	
37	6947	13	=1-(F$2/(F$2+J49))^F$1	=D49-D48	=C49*LN(E49)	=B$1*D49	=J48+K49	=EXP(SOMMEPROD(I$4;L49))	1	=C49	=B$1*E49	
38Q	5733	51	=1-(F$2/(F$2+J50))^F$1	=D50-D49	=C50*LN(E50)	=B$1*D50	=J49+K50	=EXP(SOMMEPROD(I$4;L50))	1	=C50	=B$1*E50	
38V	1201	111	=1-(F$2/(F$2+J51))^F$1	=D51-D49	=C51*LN(E51)	=B$1*D51	=J50+K51	=EXP(SOMMEPROD(I$4;L51))	2	=C51	=B$1*E51	
39	6772	593	=1-(F$2/(F$2+J52))^F$1	=D52-D51	=C52*LN(E52)	=B$1*D52	=J51+K52	=EXP(SOMMEPROD(I$4;L52))	1	=C52	=B$1*E52	
40	6179	58	=1-(F$2/(F$2+J53))^F$1	=D53-D52	=C53*LN(E53)	=B$1*D53	=J52+K53	=EXP(SOMMEPROD(I$4;L53))	2	=C53	=B$1*E53	
41	6121	47	=1-(F$2/(F$2+J54))^F$1	=D54-D53	=C54*LN(E54)	=B$1*D54	=J53+K54	=EXP(SOMMEPROD(I$4;L54))	1	=C54	=B$1*E54	

Tableau 5.35 : Modèle EG + 1 covarié (type de question) : Résultats

nbre total de joueurs	227879			r	3,8289						
				alpha	475,6871						
				B_no	2,0218						
				LL=	-592815,7088		2,0218				
Questions	Nbre envois	Nbre non réponses	P(T<=t)	¦Non repons	LL	E(T(t))	A(t)	exp(BX)	Covar. type	PAR Observed	QUESTION Expected
1	227879	13111	0,0585	0,0585	-37212,4005	13337,1956	7,5521	7,5521	1	13111	13337,196
2	214768	13146	0,1128	0,0543	-38303,9491	25704,8017	15,1041	7,5521	1	13146	12367,606
3	201622	14729	0,1632	0,0504	-44010,9677	37186,6364	22,6562	7,5521	1	14729	11481,835
4	186893	59518	0,4474	0,2842	-74884,4233	101943,1101	79,6897	57,0335	2	59518	64756,474
5	127375	5216	0,4752	0,0279	-18677,4709	108290,2945	87,2418	7,5521	1	5216	6347,1844
6	122159	1896	0,5013	0,0261	-6912,0437	114239,2886	94,7938	7,5521	1	1896	5948,994
7	120263	4546	0,5258	0,0245	-16863,4960	119819,8580	102,3459	7,5521	1	4546	5580,5694
8	115717	38349	0,6693	0,1435	-74462,9145	152510,7839	159,3794	57,0335	2	38349	32690,926
9	77368	2431	0,6839	0,0146	-10269,1033	155846,1576	166,9315	7,5521	1	2431	3335,3737
10	74937	2810	0,6977	0,0138	-12029,5637	158997,5087	174,4835	7,5521	1	2810	3151,3511
11	72127	21460	0,7809	0,0832	-53359,1539	177958,1956	231,5171	57,0335	2	21460	18960,687
12	50667	1305	0,7897	0,0087	-6186,8573	179947,7844	239,0691	7,5521	1	1305	1989,5888
13Q	24551	1679	0,7980	0,0083	-8045,6196	181838,4013	246,6212	7,5521	1	1679	1890,6169
13V	24811	5457	0,8490	0,0510	-16238,6054	193462,4512	303,6548	57,0335	2	5457	11624,05
14	42226	1498	0,8544	0,0055	-7800,9343	194710,0846	311,2068	7,5521	1	1498	1247,6334
15Q	25501	1833	0,8597	0,0052	-9630,4223	195901,2140	318,7589	7,5521	1	1833	1191,1293
15V	15227	4228	0,8924	0,0327	-14459,1962	203356,6851	375,7924	57,0335	2	4228	7455,4711
16	34667	7488	0,9160	0,0237	-28036,1292	208747,3762	432,8260	57,0335	2	7488	5390,6912
17Q	18325	531	0,9187	0,0026	-3156,7088	209344,2659	440,3780	7,5521	1	531	596,8897
17V	8854	2048	0,9355	0,0168	-8369,7585	213171,0266	497,4116	57,0335	2	2048	3826,7607
18Q	7273	314	0,9373	0,0019	-1970,3972	213600,0106	504,9636	7,5521	1	314	428,98403
18V	17327	3462	0,9495	0,0122	-15256,0008	216379,0720	561,9972	57,0335	2	3462	2779,0613
19	20824	719	0,9509	0,0014	-4734,1169	216693,9744	569,5492	7,5521	1	719	314,90242
20	20105	3692	0,9600	0,0090	-17377,1188	218752,7667	626,5828	57,0335	2	3692	2058,7923
21	16413	287	0,9610	0,0010	-1973,0837	218988,2674	634,1348	7,5521	1	287	235,50074
22	16126	329	0,9620	0,0010	-2272,6386	219216,1554	641,6869	7,5521	1	329	227,88791
23Q	4113	196	0,9630	0,0010	-1360,3094	219436,7255	649,2389	7,5521	1	196	220,57016
23V	11624	1710	0,9693	0,0064	-8640,3594	220893,1334	706,2725	57,0335	2	1710	1456,4078
24	13831	1738	0,9744	0,0051	-9187,3472	222046,4642	763,3060	57,0335	2	1738	1153,3308
25	12093	214	0,9750	0,0006	-1591,6628	222180,6055	770,8581	7,5521	1	214	134,14128
26Q	10130	177	0,9756	0,0006	-1321,6470	222310,8792	778,4101	7,5521	1	177	130,27371
26V	1749	229	0,9794	0,0038	-1274,6074	223182,7132	835,4437	57,0335	2	229	871,83404
27	11473	247	0,9798	0,0004	-1904,4057	223284,8615	842,9958	7,5521	1	247	102,14829
28	11226	1475	0,9829	0,0030	-8560,0455	223972,4217	900,0293	57,0335	2	1475	687,5602
29	9751	99	0,9832	0,0004	-786,2275	224053,4565	907,5814	7,5521	1	99	81,034762
30	9652	80	0,9836	0,0003	-637,4445	224132,3827	915,1334	7,5521	1	80	78,926273
31	9572	101	0,9839	0,0003	-807,4220	224209,2664	922,6855	7,5521	1	101	76,883651
32	9471	951	0,9862	0,0023	-5783,1816	224730,0902	979,7190	57,0335	2	951	520,82376
33	8520	368	0,9865	0,0003	-3022,3427	224791,8765	987,2711	7,5521	1	368	61,786343
34	8152	177	0,9867	0,0003	-1458,0934	224852,1417	994,8231	7,5521	1	177	60,265176
35	7975	120	0,9870	0,0003	-991,5139	224910,9307	1002,3752	7,5521	1	120	58,788986
36	7855	908	0,9887	0,0018	-5760,1483	225311,4630	1059,4087	57,0335	2	908	400,5323
37	6947	13	0,9889	0,0002	-110,1056	225359,2575	1066,9608	7,5521	1	13	47,794514
38Q	5733	51	0,9891	0,0002	-433,1583	225405,9354	1074,5128	7,5521	1	51	46,677894
38V	1201	111	0,9905	0,0014	-729,2451	225725,4471	1131,5464	57,0335	2	111	319,51172
39	6772	593	0,9907	0,0002	-5153,6915	225763,7563	1139,0985	7,5521	1	593	38,309219
40	6179	58	0,9919	0,0012	-392,2836	226026,9946	1196,1320	57,0335	2	58	263,23827
41	6121	47	0,9920	0,0001	-417,3925	226058,6807	1203,6841	7,5521	1	47	31,686098





On constate que le modèle EG+1 covarié représente bien les pics de non réponses des questions virales, ce qui est logique puisque le modèle intègre le type de question comme covarié. La variable type de question est donc un facteur explicatif de la probabilité de non réponse par question. Ce modèle fournit une représentation du phénomène avec 3 paramètres (les paramètres α et r de la loi Gamma suivie par le paramètre λ du modèle exponentiel, puis le coefficient β (noté B_no sous excel) qui est l'impact du covarié sur la probabilité de réponse, selon l'équation 5.1). Cependant, on constate des écarts entre probabilité fournie par le modèle et probabilité observée.

Nous avons testé plusieurs possibilités de variable covariée : le type de question (virale ou de qualification), le numéro de la question (de 1 à 41), la complexité de la question nombre d'items de réponses), le nombre de questions virales déjà répondues. Lorsqu'on compare ces modèles avec 1 covarié, on obtient les résultats suivants :

Tableau 5.36: Comparaison des modèles Exponentiel Gamma à 1 covarié

	LL	Nbre de paramètres
EG + 1 covarié (type)	-592816	3
EG + 1 covarié (numéro)	-671896	3
EG + 1 covarié (complexité)	-610703	3
EG + 1 covarié (nbre de questions virales avant)	-661324	3

Ce tableau montre que la variable covariée permettant d'avoir le modèle le plus proche des données observées est le type de question (le modèle EG + 1 covarié (type) donne la meilleure vraisemblance). Les autres variables donnent des niveaux de vraisemblances moins élevés.
Cette comparaison justifie notre choix de prendre comme première variable covariée le type de question.

Modèle EG + 2 covariés:

Le modèle présentant encore quelques écarts avec les données observées, on a donc intégré un deuxième covarié. Plusieurs variables covariées ont été testées :

- La complexité de la question (nombre de choix possible de réponses)
- Le numéro de la question
- Le nombre de questions virales avant

Le tableau 5.37 donne une comparaison de la vraisemblance obtenue avec ces différents modèles (le test du chi-2 n'est ici pas fourni puisque ces modèles mont tous le même nombre de paramètres).

Tableau 5.37: Comparaison des modèles Exponentiel Gamma à 2 covariés

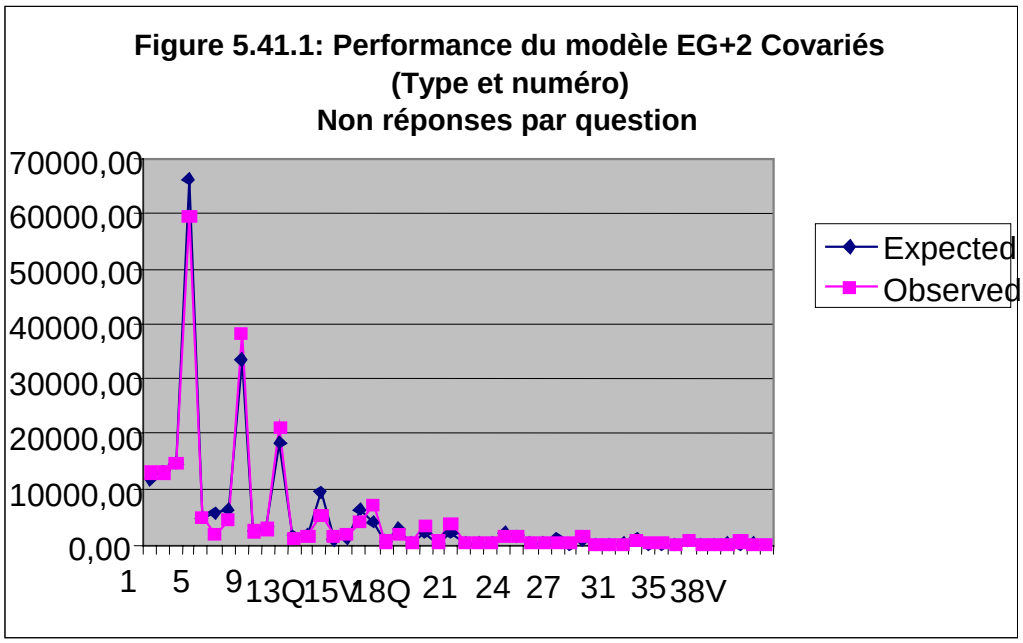
	LL	Nbre de paramètres
EG + 2 covariés (type + complexité)	-592479	4
EG + 2 covariés (type + numéro)	-589968	4
EG + 2 covariés (type + nbre de questions virales avant)	-590534	4

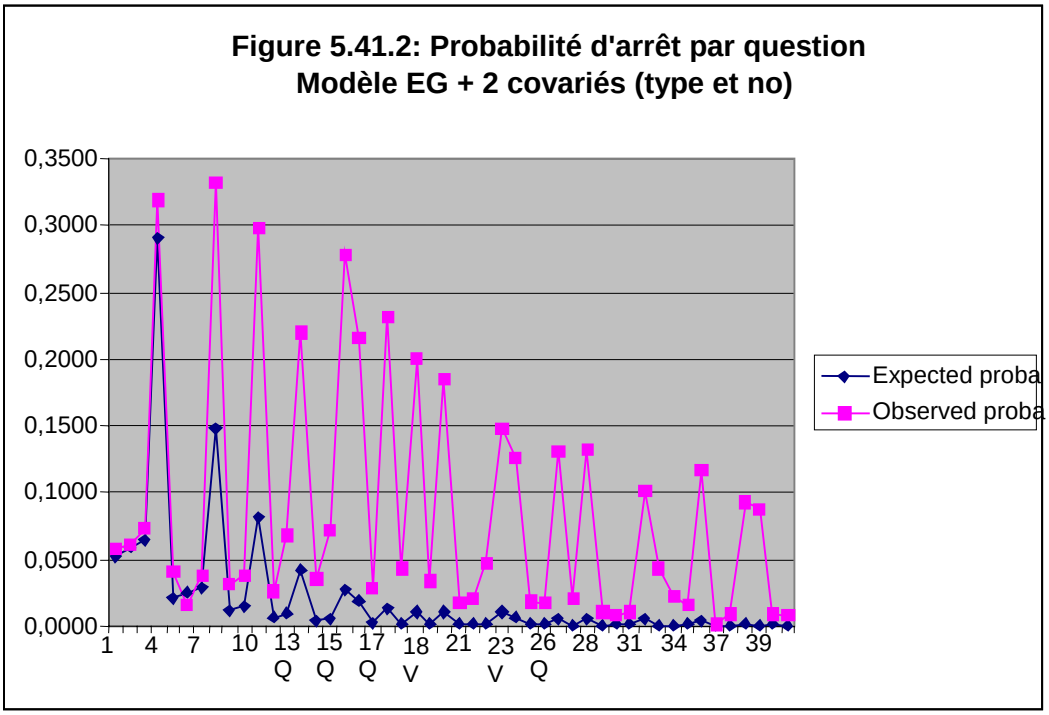
On constate que le 2ème covarié qui donne la meilleure vraisemblance est le numéro de la question, qui semble plus complémentaire de la première variable covariée (type) pour expliquer la probabilité de réponse aux questions (il est clair que la complexité est liée au type de question : pour une question virale, la complexité vaut 1 alors qu'elle correspond au nombre de réponses possibles pour une question de qualification).
Nous retiendrons donc comme 2ème covarié le numéro de la question.
Pour vérifier si ce 2ème covarié rend le modèle plus performant de manière significative, nous menons un test du ratio de vraisemblance :

Comparaison des modèles

		nbr paramètres
LL Exponentiel Gamma + 1 covarié (type)	-592816	3
LL Exponentiel Gamma + 2 Covariés (type, numéro)	-589968	4
LR(df=1)	5695,20	
Chi-2 (0,95,1)	3,841	

Etant donné que LR est nettement supérieur à 3,84, on peut conclure que le modèle avec 2 covariés (type et numéro) est meilleur que le modèle à 1 covarié (type).
La performance du modèle (par question) est fournie dans les figure 5.41.1 et 5.41.2 (le premier graphique fournit le nombre brut de non réponses, le deuxième la probabilité d'arrêt par question) , les tableaux Excel des formules et des résultats se trouvant sur la page suivante. Ce modèle inclut comme le précédent l'hétérogénéité entre les individus puisqu'on fait toujours l'hypothèse que le paramètre λ suit une loi Gamma. On constate une bonne performance du modèle sur les premières questions puis un décalage entre les probabilités observées et prévues par le modèle.





On constate une bonne prévision de probabilité pour les premières questions puis des décalages sur les questions suivantes.

Tableau 5.38 : Modèle EG + 2 covariés (type + numéro) : Formules

nbre total de joueurs	227879				r	0,370227154875036						
					alpha	78,5002645828887						
					B_type	2,14928363239204						
					B_no	0,344680454903732						
					LL=	=SOMME(F8:F55)		=TR	=TRANSPOS			
Questic	Nbre cas	Nbre non réponses	P(T<=t)	P(non repons	LL	E(T(t))	A(t)	exp(BX)		Covaria type	no	REPONSE Observed
1	227879	13111	=1-(F$2/(F$2+J8))^F$1	=D8	=C8*LN(E8)	=B$1*D8	=K8	=EXP(SOMMEPROD(I$5:J$5;L8:M8))		1	1	=C8
2	214768	13146	=1-(F$2/(F$2+J9))^F$1	=D9-D8	=C9*LN(E9)	=B$1*D9	=J8+K9	=EXP(SOMMEPROD(I$5:J$5;L9:M9))		1	2	=C9
3	201622	14729	=1-(F$2/(F$2+J10))^F$1	=D10-D9	=C10*LN(E10)	=B$1*D10	=J9+K10	=EXP(SOMMEPROD(I$5:J$5;L10:M10))		1	3	=C10
4	186893	59518	=1-(F$2/(F$2+J11))^F$1	=D11-D10	=C11*LN(E11)	=B$1*D11	=J10+K11	=EXP(SOMMEPROD(I$5:J$5;L11:M11))		2	4	=C11
5	127375	5216	=1-(F$2/(F$2+J12))^F$1	=D12-D11	=C12*LN(E12)	=B$1*D12	=J11+K12	=EXP(SOMMEPROD(I$5:J$5;L12:M12))		1	5	=C12
6	122159	1896	=1-(F$2/(F$2+J13))^F$1	=D13-D12	=C13*LN(E13)	=B$1*D13	=J12+K13	=EXP(SOMMEPROD(I$5:J$5;L13:M13))		1	6	=C13
7	120263	4546	=1-(F$2/(F$2+J14))^F$1	=D14-D13	=C14*LN(E14)	=B$1*D14	=J13+K14	=EXP(SOMMEPROD(I$5:J$5;L14:M14))		1	7	=C14
8	115717	38349	=1-(F$2/(F$2+J15))^F$1	=D15-D14	=C15*LN(E15)	=B$1*D15	=J14+K15	=EXP(SOMMEPROD(I$5:J$5;L15:M15))		2	8	=C15
9	77368	2431	=1-(F$2/(F$2+J16))^F$1	=D16-D15	=C16*LN(E16)	=B$1*D16	=J15+K16	=EXP(SOMMEPROD(I$5:J$5;L16:M16))		1	9	=C16
10	74937	2810	=1-(F$2/(F$2+J17))^F$1	=D17-D16	=C17*LN(E17)	=B$1*D17	=J16+K17	=EXP(SOMMEPROD(I$5:J$5;L17:M17))		1	10	=C17
11	72127	21460	=1-(F$2/(F$2+J18))^F$1	=D18-D17	=C18*LN(E18)	=B$1*D18	=J17+K18	=EXP(SOMMEPROD(I$5:J$5;L18:M18))		2	11	=C18
12	50667	1305	=1-(F$2/(F$2+J19))^F$1	=D19-D18	=C19*LN(E19)	=B$1*D19	=J18+K19	=EXP(SOMMEPROD(I$5:J$5;L19:M19))		1	12	=C19
13Q	24551	1679	=1-(F$2/(F$2+J20))^F$1	=D20-D19	=C20*LN(E20)	=B$1*D20	=J19+K20	=EXP(SOMMEPROD(I$5:J$5;L20:M20))		1	13	=C20
13V	24811	5457	=1-(F$2/(F$2+J21))^F$1	=D21-D20	=C21*LN(E21)	=B$1*D21	=J20+K21	=EXP(SOMMEPROD(I$5:J$5;L21:M21))		2	13	=C21
14	42226	1498	=1-(F$2/(F$2+J22))^F$1	=D22-D21	=C22*LN(E22)	=B$1*D22	=J21+K22	=EXP(SOMMEPROD(I$5:J$5;L22:M22))		1	14	=C22
15Q	25501	1833	=1-(F$2/(F$2+J23))^F$1	=D23-D22	=C23*LN(E23)	=B$1*D23	=J22+K23	=EXP(SOMMEPROD(I$5:J$5;L23:M23))		1	15	=C23
15V	15227	4228	=1-(F$2/(F$2+J24))^F$1	=D24-D23	=C24*LN(E24)	=B$1*D24	=J23+K24	=EXP(SOMMEPROD(I$5:J$5;L24:M24))		2	15	=C24
16	34667	7488	=1-(F$2/(F$2+J25))^F$1	=D25-D24	=C25*LN(E25)	=B$1*D25	=J24+K25	=EXP(SOMMEPROD(I$5:J$5;L25:M25))		2	16	=C25
17Q	18325	531	=1-(F$2/(F$2+J26))^F$1	=D26-D25	=C26*LN(E26)	=B$1*D26	=J25+K26	=EXP(SOMMEPROD(I$5:J$5;L26:M26))		1	17	=C26
17V	8854	2048	=1-(F$2/(F$2+J27))^F$1	=D27-D26	=C27*LN(E27)	=B$1*D27	=J26+K27	=EXP(SOMMEPROD(I$5:J$5;L27:M27))		2	17	=C27
18Q	7273	314	=1-(F$2/(F$2+J28))^F$1	=D28-D27	=C28*LN(E28)	=B$1*D28	=J27+K28	=EXP(SOMMEPROD(I$5:J$5;L28:M28))		1	18	=C28
18V	17327	3462	=1-(F$2/(F$2+J29))^F$1	=D29-D28	=C29*LN(E29)	=B$1*D29	=J28+K29	=EXP(SOMMEPROD(I$5:J$5;L29:M29))		2	18	=C29
19	20824	719	=1-(F$2/(F$2+J30))^F$1	=D30-D29	=C30*LN(E30)	=B$1*D30	=J29+K30	=EXP(SOMMEPROD(I$5:J$5;L30:M30))		1	19	=C30
20	20105	3692	=1-(F$2/(F$2+J31))^F$1	=D31-D30	=C31*LN(E31)	=B$1*D31	=J30+K31	=EXP(SOMMEPROD(I$5:J$5;L31:M31))		2	20	=C31
21	16413	287	=1-(F$2/(F$2+J32))^F$1	=D32-D31	=C32*LN(E32)	=B$1*D32	=J31+K32	=EXP(SOMMEPROD(I$5:J$5;L32:M32))		1	21	=C32
22	16126	329	=1-(F$2/(F$2+J33))^F$1	=D33-D32	=C33*LN(E33)	=B$1*D33	=J32+K33	=EXP(SOMMEPROD(I$5:J$5;L33:M33))		1	22	=C33
23Q	4113	196	=1-(F$2/(F$2+J34))^F$1	=D34-D33	=C34*LN(E34)	=B$1*D34	=J33+K34	=EXP(SOMMEPROD(I$5:J$5;L34:M34))		1	23	=C34
23V	11624	1710	=1-(F$2/(F$2+J35))^F$1	=D35-D34	=C35*LN(E35)	=B$1*D35	=J34+K35	=EXP(SOMMEPROD(I$5:J$5;L35:M35))		2	23	=C35
24	13831	1738	=1-(F$2/(F$2+J36))^F$1	=D36-D35	=C36*LN(E36)	=B$1*D36	=J35+K36	=EXP(SOMMEPROD(I$5:J$5;L36:M36))		2	24	=C36
25	12093	214	=1-(F$2/(F$2+J37))^F$1	=D37-D36	=C37*LN(E37)	=B$1*D37	=J36+K37	=EXP(SOMMEPROD(I$5:J$5;L37:M37))		1	25	=C37
26Q	10130	177	=1-(F$2/(F$2+J38))^F$1	=D38-D37	=C38*LN(E38)	=B$1*D38	=J37+K38	=EXP(SOMMEPROD(I$5:J$5;L38:M38))		1	26	=C38
26V	1749	229	=1-(F$2/(F$2+J39))^F$1	=D39-D38	=C39*LN(E39)	=B$1*D39	=J38+K39	=EXP(SOMMEPROD(I$5:J$5;L39:M39))		2	26	=C39
27	11473	247	=1-(F$2/(F$2+J40))^F$1	=D40-D39	=C40*LN(E40)	=B$1*D40	=J39+K40	=EXP(SOMMEPROD(I$5:J$5;L40:M40))		1	27	=C40
28	11226	1475	=1-(F$2/(F$2+J41))^F$1	=D41-D40	=C41*LN(E41)	=B$1*D41	=J40+K41	=EXP(SOMMEPROD(I$5:J$5;L41:M41))		2	28	=C41
29	9751	99	=1-(F$2/(F$2+J42))^F$1	=D42-D41	=C42*LN(E42)	=B$1*D42	=J41+K42	=EXP(SOMMEPROD(I$5:J$5;L42:M42))		1	29	=C42
30	9652	80	=1-(F$2/(F$2+J43))^F$1	=D43-D42	=C43*LN(E43)	=B$1*D43	=J42+K43	=EXP(SOMMEPROD(I$5:J$5;L43:M43))		1	30	=C43
31	9572	101	=1-(F$2/(F$2+J44))^F$1	=D44-D43	=C44*LN(E44)	=B$1*D44	=J43+K44	=EXP(SOMMEPROD(I$5:J$5;L44:M44))		1	31	=C44
32	9471	951	=1-(F$2/(F$2+J45))^F$1	=D45-D44	=C45*LN(E45)	=B$1*D45	=J44+K45	=EXP(SOMMEPROD(I$5:J$5;L45:M45))		2	32	=C45
33	8520	368	=1-(F$2/(F$2+J46))^F$1	=D46-D45	=C46*LN(E46)	=B$1*D46	=J45+K46	=EXP(SOMMEPROD(I$5:J$5;L46:M46))		1	33	=C46
34	8152	177	=1-(F$2/(F$2+J47))^F$1	=D47-D46	=C47*LN(E47)	=B$1*D47	=J46+K47	=EXP(SOMMEPROD(I$5:J$5;L47:M47))		1	34	=C47
35	7975	120	=1-(F$2/(F$2+J48))^F$1	=D48-D47	=C48*LN(E48)	=B$1*D48	=J47+K48	=EXP(SOMMEPROD(I$5:J$5;L48:M48))		1	35	=C48
36	7855	908	=1-(F$2/(F$2+J49))^F$1	=D49-D48	=C49*LN(E49)	=B$1*D49	=J48+K49	=EXP(SOMMEPROD(I$5:J$5;L49:M49))		2	36	=C49
37	6947	13	=1-(F$2/(F$2+J50))^F$1	=D50-D49	=C50*LN(E50)	=B$1*D50	=J49+K50	=EXP(SOMMEPROD(I$5:J$5;L50:M50))		1	37	=C50
38Q	5733	51	=1-(F$2/(F$2+J51))^F$1	=D51-D50	=C51*LN(E51)	=B$1*D51	=J50+K51	=EXP(SOMMEPROD(I$5:J$5;L51:M51))		1	38	=C51
38V	1201	111	=1-(F$2/(F$2+J52))^F$1	=D52-D51	=C52*LN(E52)	=B$1*D52	=J51+K52	=EXP(SOMMEPROD(I$5:J$5;L52:M52))		2	38	=C52
39	6772	593	=1-(F$2/(F$2+J53))^F$1	=D53-D52	=C53*LN(E53)	=B$1*D53	=J52+K53	=EXP(SOMMEPROD(I$5:J$5;L53:M53))		1	39	=C53
40	6179	58	=1-(F$2/(F$2+J54))^F$1	=D54-D53	=C54*LN(E54)	=B$1*D54	=J53+K54	=EXP(SOMMEPROD(I$5:J$5;L54:M54))		2	40	=C54
41	6121	47	=1-(F$2/(F$2+J55))^F$1	=D55-D54	=C55*LN(E55)	=B$1*D55	=J54+K55	=EXP(SOMMEPROD(I$5:J$5;L55:M55))		1	41	=C55

Tableau 5.39 : Modèle EG + 2 covariés (type + numéro) : Résultats

nbre total de joueurs	227879	r	0,3702		
		alpha	78,5003		
		B_type	2,1493		
		B_no	0,3447		
		LL=	-589968,1070	2,1493	0,3447

Questions	Nbre cas	Nbre non réponses	P(T<=t)	P(non repons	LL	E(T(t))	A(t)	exp(BX)	Covariates type	Covariates no	REPONSE QUESTION Observed	REPONSE QUESTION Expected
1	227879	13111	0,05172571	0,0517	-38832,1629	11787,2041	12,1092	12,1092	1	1	13111	11787,20
2	214768	13146	0,11049524	0,0588	-37257,4972	25179,5447	29,2018	17,0926	1	2	13146	13392,34
3	201622	14729	0,17463267	0,0641	-40456,5448	39795,1176	53,3286	24,1268	1	3	14729	14615,57
4	186893	59518	0,4644301	0,2898	-73717,3952	105833,8657	345,4850	292,1563	2	4	59518	66038,75
5	127375	5216	0,48530785	0,0209	-20181,0748	110591,4674	393,5563	48,0713	1	5	5216	4757,60
6	122159	1896	0,51027435	0,0250	-6996,6578	116280,8083	461,4108	67,8545	1	6	1896	5689,34
7	120263	4546	0,53900621	0,0287	-16137,1579	122828,1952	557,1901	95,7793	1	7	4546	6547,39
8	115717	38349	0,6861335	0,1471	-73494,2135	156355,4164	1716,9995	1159,8093	2	8	38349	33527,22
9	77368	2431	0,69765402	0,0115	-10851,0736	158980,7010	1907,8342	190,8347	1	9	2431	2625,28
10	74937	2810	0,71155928	0,0139	-12014,1221	162149,4171	2177,2048	269,3706	1	10	2810	3168,72
11	72127	21460	0,79287321	0,0813	-53852,5389	180679,1534	5439,0629	3261,8581	2	11	21460	18529,74
12	50667	1305	0,79987064	0,0070	-6475,6869	182273,7210	5975,7681	536,7053	1	12	1305	1594,57
13Q	24551	1679	0,80841833	0,0085	-7995,5554	184221,5610	6733,3484	757,5803	1	13	1679	1947,84
13V	24811	5457	0,85050074	0,0421	-17288,4602	193811,2590	13232,4107	6499,0623	2	13	5457	9589,70
14	42226	1498	0,85471709	0,0042	-8192,2429	194772,0751	14301,7647	1069,3540	1	14	1498	960,82
15Q	25501	1833	0,85998788	0,0053	-9615,1385	195973,1778	15811,1994	1509,4347	1	15	1833	1201,10
15V	15227	4228	0,88771322	0,0277	-15159,1066	202291,2012	28760,2025	12949,0032	2	15	4228	6318,02
16	34667	7488	0,90637401	0,0187	-29812,2059	206543,6027	47038,2229	18278,0204	2	16	7488	4252,40
17Q	18325	531	0,90849442	0,0021	-3268,9141	207026,7989	50045,6835	3007,4607	1	17	531	483,20
17V	8854	2048	0,92153369	0,0130	-8887,8880	209998,1767	75845,8184	25800,1349	2	17	2048	2971,38
18Q	7273	314	0,92309839	0,0016	-2028,4606	210354,7374	80090,9658	4245,1474	1	18	314	356,56
18V	17327	3462	0,93305459	0,0100	-15958,2942	212623,5475	116508,8571	36417,8913	2	18	3462	2268,81
19	20824	719	0,93428535	0,0012	-4817,3920	212904,0106	122501,0474	5992,1903	1	19	719	280,46
20	20105	3692	0,94467737	0,0104	-16860,3185	215272,1343	195061,5755	72560,5281	2	20	3692	2368,12
21	16413	287	0,94588039	0,0012	-1929,4788	215546,2766	207000,6657	11939,0902	1	21	287	274,14
22	16126	329	0,94742555	0,0015	-2129,4930	215898,3877	223853,1547	16852,4890	1	22	329	352,11
23Q	4113	196	0,94935435	0,0019	-1225,1683	216337,9201	247641,0967	23787,9421	1	23	196	439,53
23V	11624	1710	0,95945644	0,0101	-7857,4716	218639,9750	451710,9695	204069,8727	2	23	1710	2302,05
24	13831	1738	0,96622323	0,0068	-8682,5770	220181,9835	739763,5122	288052,5427	2	24	1738	1542,01
25	12093	214	0,96699086	0,0008	-1534,8503	220356,9111	787159,6017	47396,0895	1	25	214	174,93
26Q	10130	177	0,96797275	0,0010	-1225,9090	220580,6614	854061,0209	66901,4192	1	26	177	223,75
26V	1749	229	0,97352246	0,0055	-1189,4280	221845,3246	1427988,9408	573927,9199	2	26	229	1264,66
27	11473	247	0,97414277	0,0006	-1824,1668	221986,6803	1522422,8896	94433,9488	1	27	247	141,36
28	11226	1475	0,97898625	0,0048	-7861,9311	223090,4066	2665940,8537	1143517,9641	2	28	1475	1103,73
29	9751	99	0,97951016	0,0005	-747,8644	223209,7953	2854094,9985	188154,1447	1	29	99	119,39
30	9652	80	0,98017411	0,0007	-585,3842	223361,0958	3119681,8736	265586,8751	1	30	80	151,30
31	9572	101	0,98098978	0,0008	-718,2616	223546,9698	3494568,0519	374886,1783	1	31	101	185,87
32	9471	951	0,98603197	0,0050	-5030,7092	224695,9787	8034133,3754	4539565,3235	2	32	951	1149,01
33	8520	368	0,98648421	0,0005	-2834,0758	224799,0352	8781072,2584	746938,8830	1	33	368	103,06
34	8152	177	0,98703985	0,0006	-1326,6824	224925,6550	9835405,4710	1054333,2126	1	34	177	126,62
35	7975	120	0,9876986	0,0007	-879,0205	225075,7697	11323637,7059	1488232,2349	1	35	120	150,11
36	7855	908	0,99135333	0,0037	-5095,4540	225908,6055	29344914,1289	18021276,4230	2	36	908	832,84
37	6947	13	0,99165606	0,0003	-105,3348	225977,5910	32310130,2064	2965216,0776	1	37	13	68,99
38Q	5733	51	0,992024	0,0004	-403,2876	226061,4360	36495647,7422	4185517,5358	1	38	51	83,85
38V	1201	111	0,99381073	0,0018	-702,3376	226468,5951	72401991,7490	35906344,0067	2	38	111	407,16
39	6772	593	0,99398789	0,0002	-5122,6087	226508,9659	78310012,1375	5908020,3886	1	39	593	40,37
40	6179	58	0,99527196	0,0013	-386,1477	226801,5790	149851311,9695	71541299,8320	2	40	58	292,61
41	6121	47	0,9954025	0,0001	-420,3619	226831,3252	161622698,6067	11771386,6372	1	41	47	29,75

Modèle EG + 3 covariés (Type de question, numéro de la question et nbre de questions virales avant)

Afin d'améliorer le modèle nous avons essayé d'introduire une 3ème variable covariée. La comparaison des modèles avec 3 covariés donne les résultats suivants :

Tableau 5.40: Comparaison des modèles Exponentiel Gamma à 3 covariés

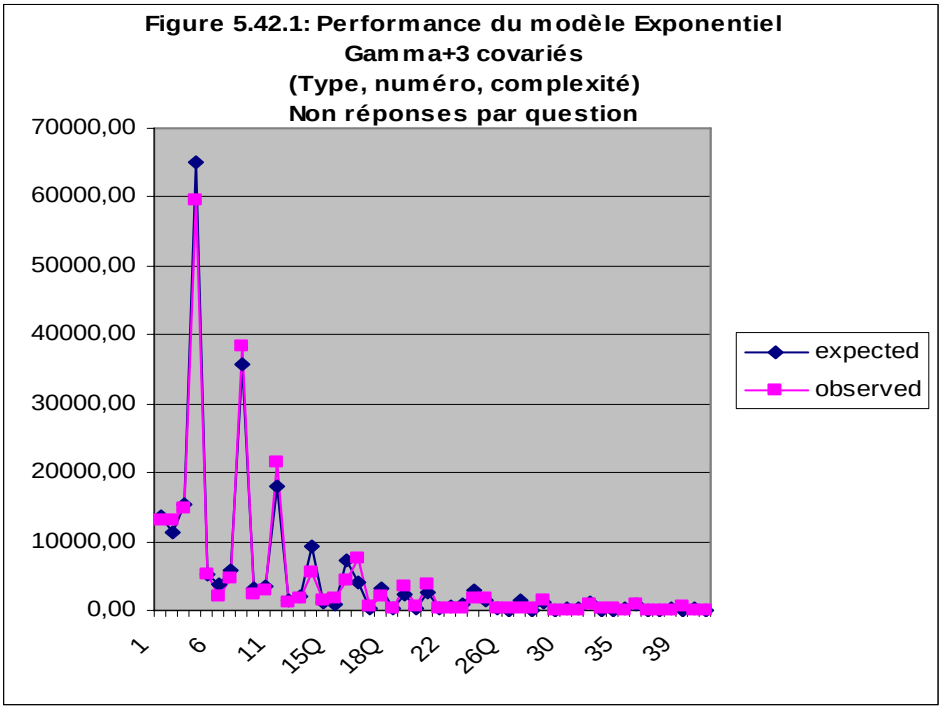
	LL	Nbre de paramètres
EG + 3 covariés (type + numéro + nb de questions virales avant)	-589946	5
EG + 3 covariés (type + numéro + complexité)	-588767	5

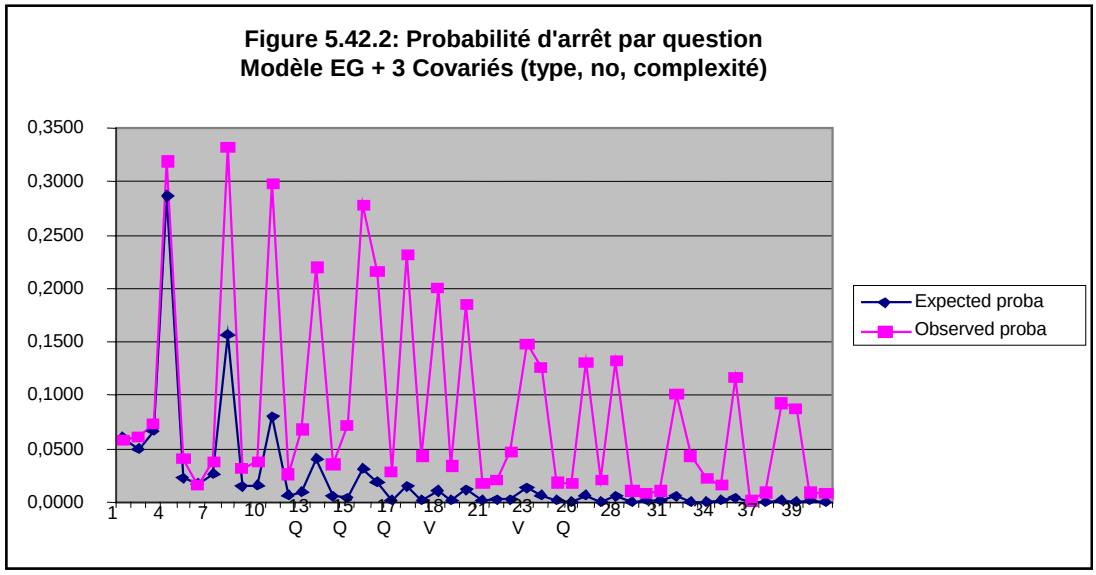
Ce tableau montre que nous devons effectivement prendre comme 3ème covarié la complexité de la question plutôt que le nombre de questions virales précédentes, puisque la vraisemblance du modèle est meilleure (-588767 versus –589946 dans le premier cas). L'apport de la variable complexité de la question est significative comme on peut le constater ci-dessous en faisant le test du ratio de vraisemblance (LR > 3,841).

Comparaison des modèles

		nbr paramètres
LL Exponentiel Gamma + 2 Covariés (type, numéro)	-589968	4
LL Exponentiel Gamma + 3 Covariés (type, numéro, complexité)	-588767	5
LR(df=1)	2402,16	
Chi-2 (0,95,1)	3,841	

On constate sur la figure 5 .42.1 que le modèle EG + 3 covariés représente correctement les données observées. La figure 5.42.2 montre que le modèle sous estime la probabilité d'arrêt à partir de la 8ème question.





Les tableaux des formules et des résultats (tableaux 5.41 et 5.42) sont fournis ci-dessous.

Tableau 5.41 : Modèle EG + 3 covariés (type, numéro, complexité) : Formules

total de joueu 227879

r	0,2816488397152
alpha	25,335644633591
B_no	0,4689903630783
B_type	1,6907391942610
B_complexité	-0,064738028392
LL=	=SOMME(F10:F5

=T =TRANSPOS =TRANSPOSE(F6)

Ques	Nbre cas	Nbre non réponses	P(T<=t)	P(non repons	LL	E(T(t))	A(t)	exp(BX)	Covariat no	type	complexit	QUESTIOI Observed
1	227879	13111	=1-(F$3/(F$3+J10))^f	=D10	=C10*LN(E10)	=B$2*D10	=K10	=EXP(SOMMEPROD(I$7:K$7;L10:N10))	1	1	5	=C10
2	214768	13146	=1-(F$3/(F$3+J11))^f	=D11-D10	=C11*LN(E11)	=B$2*D11	=J10+K11	=EXP(SOMMEPROD(I$7:K$7;L11:N11))	2	1	11	=C11
3	201622	14729	=1-(F$3/(F$3+J12))^f	=D12-D11	=C12*LN(E12)	=B$2*D12	=J11+K12	=EXP(SOMMEPROD(I$7:K$7;L12:N12))	3	1	9	=C12
4	186893	59518	=1-(F$3/(F$3+J13))^f	=D13-D12	=C13*LN(E13)	=B$2*D13	=J12+K13	=EXP(SOMMEPROD(I$7:K$7;L13:N13))	4	2	1	=C13
5	127375	5216	=1-(F$3/(F$3+J14))^f	=D14-D13	=C14*LN(E14)	=B$2*D14	=J13+K14	=EXP(SOMMEPROD(I$7:K$7;L14:N14))	5	1	6	=C14
6	122159	1896	=1-(F$3/(F$3+J15))^f	=D15-D14	=C15*LN(E15)	=B$2*D15	=J14+K15	=EXP(SOMMEPROD(I$7:K$7;L15:N15))	6	1	15	=C15
7	120263	4546	=1-(F$3/(F$3+J16))^f	=D16-D15	=C16*LN(E16)	=B$2*D16	=J15+K16	=EXP(SOMMEPROD(I$7:K$7;L16:N16))	7	1	13	=C16
8	115717	38349	=1-(F$3/(F$3+J17))^f	=D17-D16	=C17*LN(E17)	=B$2*D17	=J16+K17	=EXP(SOMMEPROD(I$7:K$7;L17:N17))	8	2	1	=C17
9	77368	2431	=1-(F$3/(F$3+J18))^f	=D18-D17	=C18*LN(E18)	=B$2*D18	=J17+K18	=EXP(SOMMEPROD(I$7:K$7;L18:N18))	9	1	5	=C18
10	74937	2810	=1-(F$3/(F$3+J19))^f	=D19-D18	=C19*LN(E19)	=B$2*D19	=J18+K19	=EXP(SOMMEPROD(I$7:K$7;L19:N19))	10	1	7	=C19
11	72127	21460	=1-(F$3/(F$3+J20))^f	=D20-D19	=C20*LN(E20)	=B$2*D20	=J19+K20	=EXP(SOMMEPROD(I$7:K$7;L20:N20))	11	2	1	=C20
12	50667	1305	=1-(F$3/(F$3+J21))^f	=D21-D20	=C21*LN(E21)	=B$2*D21	=J20+K21	=EXP(SOMMEPROD(I$7:K$7;L21:N21))	12	1	9	=C21
13Q	24551	1679	=1-(F$3/(F$3+J22))^f	=D22-D21	=C22*LN(E22)	=B$2*D22	=J21+K22	=EXP(SOMMEPROD(I$7:K$7;L22:N22))	13	1	9	=C22
13V	24811	5457	=1-(F$3/(F$3+J23))^f	=D23-D22	=C23*LN(E23)	=B$2*D23	=J22+K23	=EXP(SOMMEPROD(I$7:K$7;L23:N23))	13	2	1	=C23
14	42226	1498	=1-(F$3/(F$3+J24))^f	=D24-D23	=C24*LN(E24)	=B$2*D24	=J23+K24	=EXP(SOMMEPROD(I$7:K$7;L24:N24))	14	1	6	=C24
15Q	25501	1833	=1-(F$3/(F$3+J25))^f	=D25-D24	=C25*LN(E25)	=B$2*D25	=J24+K25	=EXP(SOMMEPROD(I$7:K$7;L25:N25))	15	1	13	=C25
15V	15227	4228	=1-(F$3/(F$3+J26))^f	=D26-D24	=C26*LN(E26)	=B$2*D26	=J25+K26	=EXP(SOMMEPROD(I$7:K$7;L26:N26))	15	2	1	=C26
16	34667	7488	=1-(F$3/(F$3+J27))^f	=D27-D26	=C27*LN(E27)	=B$2*D27	=J26+K27	=EXP(SOMMEPROD(I$7:K$7;L27:N27))	16	2	1	=C27
17Q	18325	531	=1-(F$3/(F$3+J28))^f	=D28-D27	=C28*LN(E28)	=B$2*D28	=J27+K28	=EXP(SOMMEPROD(I$7:K$7;L28:N28))	17	1	14	=C28
17V	8854	2048	=1-(F$3/(F$3+J29))^f	=D29-D27	=C29*LN(E29)	=B$2*D29	=J28+K29	=EXP(SOMMEPROD(I$7:K$7;L29:N29))	17	2	1	=C29
18Q	7273	314	=1-(F$3/(F$3+J30))^f	=D30-D29	=C30*LN(E30)	=B$2*D30	=J29+K30	=EXP(SOMMEPROD(I$7:K$7;L30:N30))	18	1	6	=C30
18V	17327	3462	=1-(F$3/(F$3+J31))^f	=D31-D30	=C31*LN(E31)	=B$2*D31	=J30+K31	=EXP(SOMMEPROD(I$7:K$7;L31:N31))	18	2	1	=C31
19	20824	719	=1-(F$3/(F$3+J32))^f	=D32-D31	=C32*LN(E32)	=B$2*D32	=J31+K32	=EXP(SOMMEPROD(I$7:K$7;L32:N32))	19	1	7	=C32
20	20105	3692	=1-(F$3/(F$3+J33))^f	=D33-D32	=C33*LN(E33)	=B$2*D33	=J32+K33	=EXP(SOMMEPROD(I$7:K$7;L33:N33))	20	2	1	=C33
21	16413	287	=1-(F$3/(F$3+J34))^f	=D34-D33	=C34*LN(E34)	=B$2*D34	=J33+K34	=EXP(SOMMEPROD(I$7:K$7;L34:N34))	21	1	11	=C34
22	16126	329	=1-(F$3/(F$3+J35))^f	=D35-D34	=C35*LN(E35)	=B$2*D35	=J34+K35	=EXP(SOMMEPROD(I$7:K$7;L35:N35))	22	1	4	=C35
23Q	4113	196	=1-(F$3/(F$3+J36))^f	=D36-D35	=C36*LN(E36)	=B$2*D36	=J35+K36	=EXP(SOMMEPROD(I$7:K$7;L36:N36))	23	1	3	=C36
23V	11624	1710	=1-(F$3/(F$3+J37))^f	=D37-D35	=C37*LN(E37)	=B$2*D37	=J36+K37	=EXP(SOMMEPROD(I$7:K$7;L37:N37))	23	2	1	=C37
24	13831	1738	=1-(F$3/(F$3+J38))^f	=D38-D37	=C38*LN(E38)	=B$2*D38	=J37+K38	=EXP(SOMMEPROD(I$7:K$7;L38:N38))	24	2	1	=C38
25	12093	214	=1-(F$3/(F$3+J39))^f	=D39-D38	=C39*LN(E39)	=B$2*D39	=J38+K39	=EXP(SOMMEPROD(I$7:K$7;L39:N39))	25	1	3	=C39
26Q	10130	177	=1-(F$3/(F$3+J40))^f	=D40-D39	=C40*LN(E40)	=B$2*D40	=J39+K40	=EXP(SOMMEPROD(I$7:K$7;L40:N40))	26	1	18	=C40
26V	1749	229	=1-(F$3/(F$3+J41))^f	=D41-D39	=C41*LN(E41)	=B$2*D41	=J40+K41	=EXP(SOMMEPROD(I$7:K$7;L41:N41))	26	2	1	=C41
27	11473	247	=1-(F$3/(F$3+J42))^f	=D42-D41	=C42*LN(E42)	=B$2*D42	=J41+K42	=EXP(SOMMEPROD(I$7:K$7;L42:N42))	27	1	15	=C42
28	11226	1475	=1-(F$3/(F$3+J43))^f	=D43-D42	=C43*LN(E43)	=B$2*D43	=J42+K43	=EXP(SOMMEPROD(I$7:K$7;L43:N43))	28	2	1	=C43
29	9751	99	=1-(F$3/(F$3+J44))^f	=D44-D43	=C44*LN(E44)	=B$2*D44	=J43+K44	=EXP(SOMMEPROD(I$7:K$7;L44:N44))	29	1	15	=C44
30	9652	80	=1-(F$3/(F$3+J45))^f	=D45-D44	=C45*LN(E45)	=B$2*D45	=J44+K45	=EXP(SOMMEPROD(I$7:K$7;L45:N45))	30	1	4	=C45
31	9572	101	=1-(F$3/(F$3+J46))^f	=D46-D45	=C46*LN(E46)	=B$2*D46	=J45+K46	=EXP(SOMMEPROD(I$7:K$7;L46:N46))	31	1	5	=C46
32	9471	951	=1-(F$3/(F$3+J47))^f	=D47-D46	=C47*LN(E47)	=B$2*D47	=J46+K47	=EXP(SOMMEPROD(I$7:K$7;L47:N47))	32	2	1	=C47
33	8520	368	=1-(F$3/(F$3+J48))^f	=D48-D47	=C48*LN(E48)	=B$2*D48	=J47+K48	=EXP(SOMMEPROD(I$7:K$7;L48:N48))	33	1	2	=C48
34	8152	177	=1-(F$3/(F$3+J49))^f	=D49-D48	=C49*LN(E49)	=B$2*D49	=J48+K49	=EXP(SOMMEPROD(I$7:K$7;L49:N49))	34	1	15	=C49
35	7975	120	=1-(F$3/(F$3+J50))^f	=D50-D49	=C50*LN(E50)	=B$2*D50	=J49+K50	=EXP(SOMMEPROD(I$7:K$7;L50:N50))	35	1	3	=C50
36	7855	908	=1-(F$3/(F$3+J51))^f	=D51-D50	=C51*LN(E51)	=B$2*D51	=J50+K51	=EXP(SOMMEPROD(I$7:K$7;L51:N51))	36	2	1	=C51
37	6947	13	=1-(F$3/(F$3+J52))^f	=D52-D51	=C52*LN(E52)	=B$2*D52	=J51+K52	=EXP(SOMMEPROD(I$7:K$7;L52:N52))	37	1	3	=C52
38Q	5733	51	=1-(F$3/(F$3+J53))^f	=D53-D52	=C53*LN(E53)	=B$2*D53	=J52+K53	=EXP(SOMMEPROD(I$7:K$7;L53:N53))	38	1	2	=C53
38V	1201	111	=1-(F$3/(F$3+J54))^f	=D54-D52	=C54*LN(E54)	=B$2*D54	=J53+K54	=EXP(SOMMEPROD(I$7:K$7;L54:N54))	38	2	1	=C54
39	6772	593	=1-(F$3/(F$3+J55))^f	=D55-D54	=C55*LN(E55)	=B$2*D55	=J54+K55	=EXP(SOMMEPROD(I$7:K$7;L55:N55))	39	1	5	=C55
40	6179	58	=1-(F$3/(F$3+J56))^f	=D56-D55	=C56*LN(E56)	=B$2*D56	=J55+K56	=EXP(SOMMEPROD(I$7:K$7;L56:N56))	40	2	1	=C56
41	6121	47	=1-(F$3/(F$3+J57))^f	=D57-D56	=C57*LN(E57)	=B$2*D57	=J56+K57	=EXP(SOMMEPROD(I$7:K$7;L57:N57))	41	1	14	=C57

Tableau 5.42 : Modèle EG + 3 covariés (type, numéro, complexité) : Résultats

total de joueurs 227879

r	0,2816
alpha	25,3356
B_no	0,4690
B_type	1,6907
B_complexit	-0,0647

LL= -587411,4345 1,6907 -0,0647

Questio	Nbre cas	Nbre non réponses	P(T<=t)	P(non repon	LL	E(T(t))	A(t)	exp(BX)	Covariates no	typ	comp	QUESTION PAR QUE Observed	Expected
1	227879	13111	0,0604	0,0604	-36800,9724	13762,3578	6,2716	6,2716	1	1	5	13111,00	13762,36
2	214768	13146	0,1106	0,0502	-39338,9311	25193,6078	13,0695	6,7978	2	1	11	13146,00	11431,25
3	201622	14729	0,1778	0,0673	-39757,4035	40519,6669	25,4369	12,3675	3	1	9	14729,00	15326,06
4	186893	59518	0,4632	0,2854	-74625,6724	105558,2784	205,3910	179,9541	4	2	1	59518,00	65038,61
5	127375	5216	0,4860	0,0228	-19730,6798	110744,9514	243,7605	38,3695	5	1	6	5216,00	5186,67
6	122159	1896	0,5030	0,0171	-7719,3125	114631,2121	278,0079	34,2474	6	1	15	1896,00	3886,26
7	120263	4546	0,5285	0,0255	-16684,6018	120435,7662	340,3151	62,3072	7	1	13	4546,00	5804,55
8	115717	38349	0,6855	0,1570	-70998,1644	156217,7817	1514,8920	1174,5769	8	2	1	38349,00	35782,02
9	77368	2431	0,6994	0,0139	-10402,6772	159374,8336	1782,0827	267,1906	9	1	5	2431,00	3157,05
10	74937	2810	0,7149	0,0156	-11699,2523	162919,2723	2157,2892	375,2065	10	1	7	2810,00	3544,44
11	72127	21460	0,7945	0,0796	-54312,9874	181055,6672	6953,7536	4796,4645	11	2	1	21460,00	18136,39
12	50667	1305	0,8010	0,0065	-6574,2310	182534,2586	7795,9231	842,1695	12	1	9	1305,00	1478,59
13Q	24551	1679	0,8097	0,0087	-7965,1372	184517,7088	9142,0297	1346,1065	13	1	9	1679,00	1983,45
13V	24811	5457	0,8502	0,0405	-17503,2430	193737,2961	21396,1202	12254,0905	13	2	1	5457,00	9219,59
14	42226	1498	0,8550	0,0048	-8004,7430	194826,2251	24008,9160	2612,7958	14	1	6	1498,00	1088,93
15Q	25501	1833	0,8592	0,0042	-10023,9308	195787,2277	26663,3859	2654,4699	15	1	13	1833,00	961,00
15V	15227	4228	0,8868	0,0319	-14570,1732	202088,5502	57970,3487	31306,9628	15	2	1	4228,00	7262,33
16	34667	7488	0,9050	0,0182	-30003,6070	206233,6333	108010,7593	50040,4106	16	2	1	7488,00	4145,08
17Q	18325	531	0,9065	0,0015	-3446,6454	206579,3821	114367,3215	6356,5622	17	1	14	531,00	345,75
17V	8854	2048	0,9195	0,0145	-8672,9855	209533,7525	194350,8927	79983,5712	17	2	1	2048,00	3300,12
18Q	7273	314	0,9214	0,0019	-1970,0724	209963,1805	211404,8500	17053,9574	18	1	6	314,00	429,43
18V	17327	3462	0,9312	0,0098	-16011,4913	212197,3945	339248,9582	127844,1081	18	2	1	3462,00	2234,21
19	20824	719	0,9326	0,0014	-4728,3779	212514,8206	364798,8892	25549,9310	19	1	7	719,00	317,43
20	20105	3692	0,9437	0,0111	-16613,4028	215046,7371	691417,2322	326618,3431	20	2	1	3692,00	2531,92
21	16413	287	0,9448	0,0011	-1953,9894	215298,4388	741800,7064	50383,4742	21	1	11	287,00	251,70
22	16126	329	0,9472	0,0024	-1984,8739	215844,9302	868500,4152	126699,7089	22	1	4	329,00	546,49
23Q	4113	196	0,9504	0,0032	-1125,7595	216574,8242	1084558,6941	216058,2789	23	1	3	196,00	729,89
23V	11624	1710	0,9604	0,0132	-7396,0135	218860,1138	2418326,8724	1333768,1782	23	2	1	1710,00	3015,18
24	13831	1738	0,9669	0,0065	-8764,7450	220330,9167	4550194,7132	2131867,8409	24	2	1	1738,00	1470,80
25	12093	214	0,9679	0,0011	-1467,5892	220570,4463	5102184,1485	551989,4353	25	1	3	214,00	239,53
26Q	10130	177	0,9685	0,0006	-1322,8350	220699,8486	5436286,9867	334102,8382	26	1	18	177,00	129,40
26V	1749	229	0,9741	0,0062	-1165,4628	221974,6318	10882819,8927	5446532,9060	26	2	1	229,00	1404,19
27	11473	247	0,9745	0,0004	-1921,0984	222070,1048	11531314,2434	648494,3508	27	1	15	247,00	95,47
28	11226	1475	0,9796	0,0051	-7787,6128	223230,8675	25446210,9024	13914896,6590	28	2	1	1475,00	1160,76
29	9751	99	0,9800	0,0004	-785,2390	223312,7154	27102995,5085	1656784,6061	29	1	15	99,00	81,85
30	9652	80	0,9810	0,0010	-552,6815	223540,4207	32500773,6579	5397778,1494	30	1	4	80,00	227,71
31	9572	101	0,9821	0,0012	-683,1308	223803,6167	40587627,4111	8086853,7532	31	1	5	101,00	263,20
32	9471	951	0,9872	0,0050	-5031,4361	224951,7477	131411432,3949	90823804,9839	32	2	1	951,00	1148,13
33	8520	368	0,9878	0,0006	-2719,8451	225092,3151	156500559,2962	25089126,9012	33	1	2	368,00	140,57
34	8152	177	0,9881	0,0004	-1405,7057	225173,3378	173785412,0581	17284852,7619	34	1	15	177,00	81,02
35	7975	120	0,9891	0,0010	-834,7635	225390,4051	233865427,1626	60080015,1045	35	1	3	120,00	217,07
36	7855	908	0,9923	0,0033	-5196,9481	226135,1628	826680721,0097	592815293,8471	36	2	1	908,00	744,76
37	6947	13	0,9927	0,0004	-103,1398	226216,8369	980174184,8107	153493463,8009	37	1	3	13,00	81,67
38Q	5733	51	0,9932	0,0005	-390,7597	226324,0281	1241923275,6862	261749090,8756	38	1	2	51,00	107,19
38V	1201	111	0,9945	0,0018	-698,9070	226636,7764	2756458062,1451	1514534786,4589	38	2	1	111,00	419,94
39	6772	593	0,9947	0,0002	-5120,2873	226677,3056	3100981683,2315	344523621,0864	39	1	5	593,00	40,53
40	6179	58	0,9958	0,0011	-396,4217	226922,4169	6970341403,2013	3869359719,9698	40	2	1	58,00	245,11
41	6121	47	0,9959	0,0001	-443,4948	226940,6003	7461860663,2182	491519260,0169	41	1	14	47,00	18,18

L'analyse des figures 5.42.1 et 5.42.2 montre qu'il reste sur certains points des écarts entre données observées et modélisées, ce qui nous a poussé à tester l'ajout d'une quatrième variable explicative possible de la probabilité de non réponse : le nombre de questions virales répondues précédemment.

Modèle EG + 4 covariés (Type, numéro, complexité de la question, nombre de questions virales précédentes) :

Le modèle avec 4 covariés intègre en plus la variable du nombre de questions virales précédentes. En appliquant la même méthode (maximisation de la vraisemblance sous Excel), on obtient pour ce modèle EG + 4 covariés les performances suivantes (les tableaux 5.43 et 5.44 détaillant les formules et les résultats).

Figure 5.43.1: Performance du modèle exponentiel gamma+ 4 covariés (Type, numéro, complexité, nb de questions virales avant) Non réponses par question

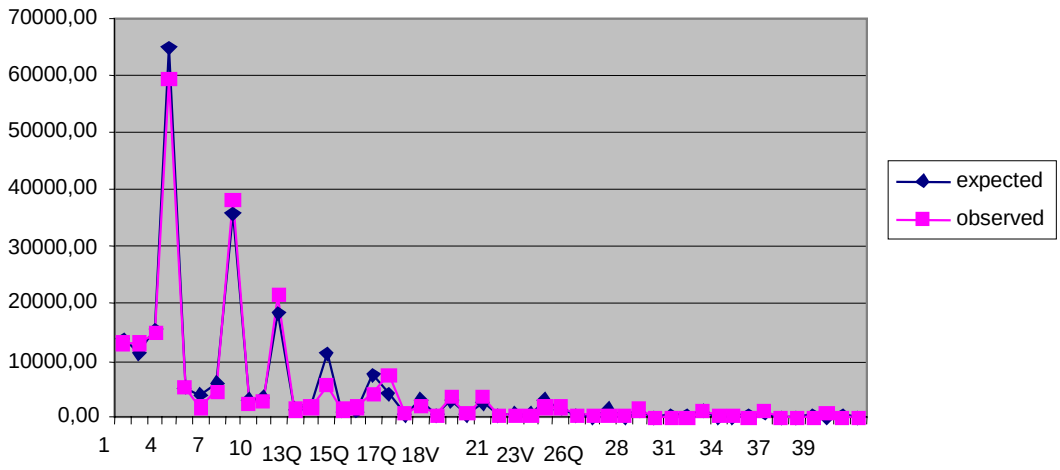


Figure 5.43.2: Probabilité d'arrêt par question Modèle EG+4covariés

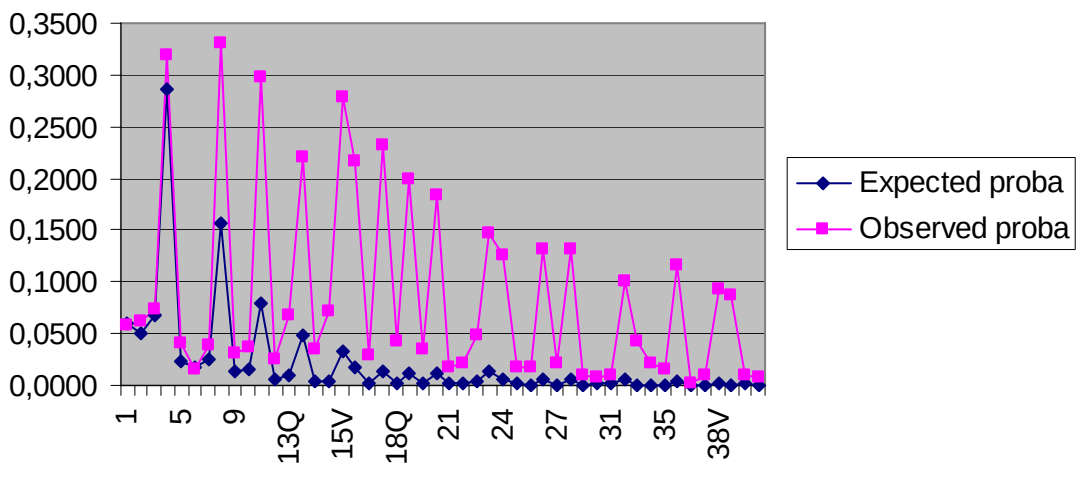


Tableau 5.43 : Modèle EG + 4 covariés : Formules

Nbre total de joueurs 227879

r	0,283502996185204
alpha	25,139171555692
B_no	0,459512253092114
B_type	1,68421466498567
B_complex	-0,064827031386858
B_qvirales	0,0088760378866322

LL= =SOMME(F12:F59)

									Covariates				QUESTION
Questions	Nbre envois	Nbre non réponse	P(T<=t)	P(non repon	LL	E(T(t))	A(t)	exp(BX)	no	type	complexité	Nbre	Observed
1	227879	13111	=1-(F$3/(F$3+J12))^F$2	=D12	=C12*LN(E12)	=B$2*D12	=K12	=EXP(SOMI	1	1	5	0	=C12
2	214768	13146	=1-(F$3/(F$3+J13))^F$2	=D13-D12	=C13*LN(E13)	=B$2*D13	=J12+K13	=EXP(SOMI	2	1	11	1	=C13
3	201622	14729	=1-(F$3/(F$3+J14))^F$2	=D14-D13	=C14*LN(E14)	=B$2*D14	=J13+K14	=EXP(SOMI	3	1	9	2	=C14
4	186893	59518	=1-(F$3/(F$3+J15))^F$2	=D15-D14	=C15*LN(E15)	=B$2*D15	=J14+K15	=EXP(SOMI	4	2	1	3	=C15
5	127375	5216	=1-(F$3/(F$3+J16))^F$2	=D16-D15	=C16*LN(E16)	=B$2*D16	=J15+K16	=EXP(SOMI	5	1	6	3	=C16
6	122159	1896	=1-(F$3/(F$3+J17))^F$2	=D17-D16	=C17*LN(E17)	=B$2*D17	=J16+K17	=EXP(SOMI	6	1	15	4	=C17
7	120263	4546	=1-(F$3/(F$3+J18))^F$2	=D18-D17	=C18*LN(E18)	=B$2*D18	=J17+K18	=EXP(SOMI	7	1	13	5	=C18
8	115717	38349	=1-(F$3/(F$3+J19))^F$2	=D19-D18	=C19*LN(E19)	=B$2*D19	=J18+K19	=EXP(SOMI	8	2	1	6	=C19
9	77368	2431	=1-(F$3/(F$3+J20))^F$2	=D20-D19	=C20*LN(E20)	=B$2*D20	=J19+K20	=EXP(SOMI	9	1	5	6	=C20
10	74937	2810	=1-(F$3/(F$3+J21))^F$2	=D21-D20	=C21*LN(E21)	=B$2*D21	=J20+K21	=EXP(SOMI	10	1	7	7	=C21
11	72127	21460	=1-(F$3/(F$3+J22))^F$2	=D22-D21	=C22*LN(E22)	=B$2*D22	=J21+K22	=EXP(SOMI	11	2	1	8	=C22
12	50667	1305	=1-(F$3/(F$3+J23))^F$2	=D23-D22	=C23*LN(E23)	=B$2*D23	=J22+K23	=EXP(SOMI	12	1	9	8	=C23
13Q	24551	1679	=1-(F$3/(F$3+J24))^F$2	=D24-D23	=C24*LN(E24)	=B$2*D24	=J23+K24	=EXP(SOMI	13	1	9	9,11	=C24
13V	24811	5457	=1-(F$3/(F$3+J25))^F$2	=D25-D23	=C25*LN(E25)	=B$2*D25	=J24+K25	=EXP(SOMI	13	2	1	9	=C25
14	42226	1498	=1-(F$3/(F$3+J26))^F$2	=D26-D25	=C26*LN(E26)	=B$2*D26	=J25+K26	=EXP(SOMI	14	1	6	9,55	=C26
15Q	25501	1833	=1-(F$3/(F$3+J27))^F$2	=D27-D26	=C27*LN(E27)	=B$2*D27	=J26+K27	=EXP(SOMI	15	1	13	10,3	=C27
15V	15227	4228	=1-(F$3/(F$3+J28))^F$2	=D28-D26	=C28*LN(E28)	=B$2*D28	=J27+K28	=EXP(SOMI	15	2	1	10,8	=C28
16	34667	7488	=1-(F$3/(F$3+J29))^F$2	=D29-D28	=C29*LN(E29)	=B$2*D29	=J28+K29	=EXP(SOMI	16	2	1	11,1	=C29
17Q	18325	531	=1-(F$3/(F$3+J30))^F$2	=D30-D29	=C30*LN(E30)	=B$2*D30	=J29+K30	=EXP(SOMI	17	1	14	11,3	=C30
17V	8854	2048	=1-(F$3/(F$3+J31))^F$2	=D31-D29	=C31*LN(E31)	=B$2*D31	=J30+K31	=EXP(SOMI	17	2	1	10,8	=C31
18Q	7273	314	=1-(F$3/(F$3+J32))^F$2	=D32-D31	=C32*LN(E32)	=B$2*D32	=J31+K32	=EXP(SOMI	18	1	6	11,8	=C32
18V	17327	3462	=1-(F$3/(F$3+J33))^F$2	=D33-D31	=C33*LN(E33)	=B$2*D33	=J32+K33	=EXP(SOMI	18	2	1	11,9	=C33
19	20824	719	=1-(F$3/(F$3+J34))^F$2	=D34-D33	=C34*LN(E34)	=B$2*D34	=J33+K34	=EXP(SOMI	19	1	7	12,2	=C34
20	20105	3692	=1-(F$3/(F$3+J35))^F$2	=D35-D34	=C35*LN(E35)	=B$2*D35	=J34+K35	=EXP(SOMI	20	2	1	13,2	=C35
21	16413	287	=1-(F$3/(F$3+J36))^F$2	=D36-D35	=C36*LN(E36)	=B$2*D36	=J35+K36	=EXP(SOMI	21	1	11	13,2	=C36
22	16126	329	=1-(F$3/(F$3+J37))^F$2	=D37-D36	=C37*LN(E37)	=B$2*D37	=J36+K37	=EXP(SOMI	22	1	4	14,2	=C37
23Q	4113	196	=1-(F$3/(F$3+J38))^F$2	=D38-D37	=C38*LN(E38)	=B$2*D38	=J37+K38	=EXP(SOMI	23	1	3	15,4	=C38
23V	11624	1710	=1-(F$3/(F$3+J39))^F$2	=D39-D37	=C39*LN(E39)	=B$2*D39	=J38+K39	=EXP(SOMI	23	2	1	15,1	=C39
24	13831	1738	=1-(F$3/(F$3+J40))^F$2	=D40-D39	=C40*LN(E40)	=B$2*D40	=J39+K40	=EXP(SOMI	24	2	1	15,5	=C40
25	12093	214	=1-(F$3/(F$3+J41))^F$2	=D41-D40	=C41*LN(E41)	=B$2*D41	=J40+K41	=EXP(SOMI	25	1	3	15,4	=C41
26Q	10130	177	=1-(F$3/(F$3+J42))^F$2	=D42-D41	=C42*LN(E42)	=B$2*D42	=J41+K42	=EXP(SOMI	26	1	18	15,4	=C42
26V	1749	229	=1-(F$3/(F$3+J43))^F$2	=D43-D41	=C43*LN(E43)	=B$2*D43	=J42+K43	=EXP(SOMI	26	2	1	16,5	=C43
27	11473	247	=1-(F$3/(F$3+J44))^F$2	=D44-D43	=C44*LN(E44)	=B$2*D44	=J43+K44	=EXP(SOMI	27	1	15	17,3	=C44
28	11226	1475	=1-(F$3/(F$3+J45))^F$2	=D45-D44	=C45*LN(E45)	=B$2*D45	=J44+K45	=EXP(SOMI	28	2	1	18,3	=C45
29	9751	99	=1-(F$3/(F$3+J46))^F$2	=D46-D45	=C46*LN(E46)	=B$2*D46	=J45+K46	=EXP(SOMI	29	1	15	18,3	=C46
30	9652	80	=1-(F$3/(F$3+J47))^F$2	=D47-D46	=C47*LN(E47)	=B$2*D47	=J46+K47	=EXP(SOMI	30	1	4	19,3	=C47
31	9572	101	=1-(F$3/(F$3+J48))^F$2	=D48-D47	=C48*LN(E48)	=B$2*D48	=J47+K48	=EXP(SOMI	31	1	5	20,3	=C48
32	9471	951	=1-(F$3/(F$3+J49))^F$2	=D49-D48	=C49*LN(E49)	=B$2*D49	=J48+K49	=EXP(SOMI	32	2	1	21,3	=C49
33	8520	368	=1-(F$3/(F$3+J50))^F$2	=D50-D49	=C50*LN(E50)	=B$2*D50	=J49+K50	=EXP(SOMI	33	1	2	21,4	=C50
34	8152	177	=1-(F$3/(F$3+J51))^F$2	=D51-D50	=C51*LN(E51)	=B$2*D51	=J50+K51	=EXP(SOMI	34	1	15	22,0	=C51
35	7975	120	=1-(F$3/(F$3+J52))^F$2	=D52-D51	=C52*LN(E52)	=B$2*D52	=J51+K52	=EXP(SOMI	35	1	3	23,0	=C52
36	7855	908	=1-(F$3/(F$3+J53))^F$2	=D53-D52	=C53*LN(E53)	=B$2*D53	=J52+K53	=EXP(SOMI	36	2	1	24,1	=C53
37	6947	13	=1-(F$3/(F$3+J54))^F$2	=D54-D53	=C54*LN(E54)	=B$2*D54	=J53+K54	=EXP(SOMI	37	1	3	24,0	=C54
38Q	5733	51	=1-(F$3/(F$3+J55))^F$2	=D55-D54	=C55*LN(E55)	=B$2*D55	=J54+K55	=EXP(SOMI	38	1	2	25,1	=C55
38V	1201	111	=1-(F$3/(F$3+J56))^F$2	=D56-D54	=C56*LN(E56)	=B$2*D56	=J55+K56	=EXP(SOMI	38	2	1	24,7	=C56
39	6772	593	=1-(F$3/(F$3+J57))^F$2	=D57-D56	=C57*LN(E57)	=B$2*D57	=J56+K57	=EXP(SOMI	39	1	5	25,9	=C57
40	6179	58	=1-(F$3/(F$3+J58))^F$2	=D58-D57	=C58*LN(E58)	=B$2*D58	=J57+K58	=EXP(SOMI	40	2	1	26,9	=C58
41	6121	47	=1-(F$3/(F$3+J59))^F$2	=D59-D58	=C59*LN(E59)	=B$2*D59	=J58+K59	=EXP(SOMI	41	1	14	26,9	=C59

Tableau 5.44 : Modèle EG + 4 covariés : Résultats

Nbre total de joueurs 227879

r	0,2835
alpha	25,1392
B_no	0,4595
B_type	1,6842
B_compl	-0,0648
B_qvirale	0,0089
LL=	-588766,7899

						0,4595	1,6842		-0,0648	0,0089				
										Covariates				QUESTIO
Questions	Nbre envo	Nbre non répon	P(T<=t)	P(non rep	LL	E(T(t))	A(t)		exp(BX)	no	type	complexité	Nbre q vira	Observed
1	227879	13111	0,0603	0,0603	-36816,3874	13746,1866	6,1693		6,1693	1	1	5	0	13111
2	214768	13146	0,1105	0,0501	-39348,0141	25169,5411	12,8487		6,6793	2	1	11	1	13146
3	201622	14729	0,1777	0,0673	-39750,2961	40502,9974	24,9954		12,1467	3	1	9	2	14729
4	186893	59518	0,4633	0,2855	-74597,3008	105572,6194	200,6078		175,6124	4	2	1	3	59518
5	127375	5216	0,4861	0,0228	-19726,2018	110763,7471	237,9246		37,3167	5	1	6	3	5216
6	122159	1896	0,5031	0,0171	-7718,8969	114650,8597	271,1856		33,2610	6	1	15	4	1896
7	120263	4546	0,5286	0,0255	-16681,5385	120459,3265	331,6726		60,4870	7	1	13	5	4546
8	115717	38349	0,6857	0,1571	-70988,1320	156250,7041	1465,0492		1133,3765	8	2	1	6	38349
9	77368	2431	0,6995	0,0139	-10401,3552	159409,4733	1722,0158		256,9666	9	1	5	6	2431
10	74937	2810	0,7151	0,0156	-11697,4000	162956,2492	2082,5837		360,5679	10	1	7	7	2810
11	72127	21460	0,7946	0,0795	-54329,3815	181078,7943	6661,6270		4579,0434	11	2	1	8	21460
12	50667	1305	0,8011	0,0065	-6575,2707	182556,2082	7462,6797		801,0527	12	1	9	8	1305
13Q	24551	1679	0,8098	0,0087	-7964,3794	184540,5540	8743,6119		1280,9322	13	1	9	9	1679
13V	24811	5457	0,8502	0,0404	-17513,5946	193742,6687	20324,8994		11581,2875	13	2	1	9	5457
14	42226	1498	0,8550	0,0048	-8000,3614	194834,7874	22798,0209		2473,1215	14	1	6	10	1498
15Q	25501	1833	0,8592	0,0042	-10022,9474	195796,3058	25303,5022		2505,4813	15	1	13	10	1833
15V	15227	4228	0,8869	0,0277	-15164,5415	202106,2128	54814,0974		29510,5952	15	2	1	11	4228
16	34667	7488	0,9051	0,0182	-30012,7274	206246,2502	101679,6341		46865,5367	16	2	1	11	7488
17Q	18325	531	0,9066	0,0015	-3447,6471	206591,3474	107615,3914		5935,7573	17	1	14	11	531
17V	8854	2048	0,9195	0,0129	-8913,8560	209525,2868	181588,0747		73972,6833	17	2	1	11	2048
18Q	7273	314	0,9213	0,0019	-1969,1567	209955,9689	197454,6534		15866,5787	18	1	6	12	314
18V	17327	3462	0,9311	0,0098	-16013,7936	212188,6977	315715,5440		118260,8906	18	2	1	12	3462
19	20824	719	0,9325	0,0014	-4727,9410	212506,3166	339333,3459		23617,8020	19	1	7	12	719
20	20105	3692	0,9436	0,0111	-16621,2643	215032,8476	639276,1581		299942,8122	20	2	1	13	3692
21	16413	287	0,9447	0,0011	-1954,7912	215283,8471	685357,6916		46081,5335	21	1	11	13	287
22	16126	329	0,9471	0,0024	-1985,3848	215829,4906	801242,4594		115884,7679	22	1	4	14	329
23Q	4113	196	0,9503	0,0032	-1125,4646	216560,4835	999167,1539		197924,6945	23	1	3	15	196
23V	11624	1710	0,9603	0,0100	-7873,6328	218840,8842	2209541,9075		1210374,7536	23	2	1	15	1710
24	13831	1738	0,9668	0,0065	-8765,5479	220311,0078	4132849,1783		1923307,2707	24	2	1	16	1738
25	12093	214	0,9678	0,0011	-1467,6492	220550,4702	4629183,6338		496334,4555	25	1	3	15	214
26Q	10130	177	0,9684	0,0006	-1324,5682	220678,6116	4926369,2123		297185,5785	26	1	18	15	177
26V	1749	229	0,9740	0,0056	-1187,7408	221952,6268	9790862,7050		4864493,4927	26	2	1	17	229
27	11473	247	0,9744	0,0004	-1919,5193	222048,7122	10371971,6718		581108,9668	27	1	15	17	247
28	11226	1475	0,9795	0,0051	-7782,2224	223213,7247	22769282,3598		12397310,6880	28	2	1	18	1475
29	9751	99	0,9799	0,0004	-785,0663	223295,7155	24238805,6358		1469523,2760	29	1	15	18	99
30	9652	80	0,9809	0,0010	-552,4328	223524,1294	29028384,0467		4789578,4109	30	1	4	19	80
31	9572	101	0,9820	0,0012	-682,7446	223788,3338	36199217,9405		7170833,8938	31	1	5	20	101
32	9471	951	0,9871	0,0051	-5028,4436	224940,0832	116209986,0612		80010768,1207	32	2	1	21	951
33	8520	368	0,9877	0,0006	-2717,9227	225081,3868	138269479,2998		22059493,2386	33	1	2	21	368
34	8152	177	0,9881	0,0004	-1405,5687	225162,4723	153384823,5095		15115344,2098	34	1	15	22	177
35	7975	120	0,9890	0,0010	-834,4036	225380,1916	205947738,1124		52562914,6028	35	1	3	23	120
36	7855	908	0,9923	0,0033	-5193,9513	226127,4114	721144680,6869		515196942,5746	36	2	1	24	908
37	6947	13	0,9927	0,0004	-103,0820	226209,4494	854092475,4536		132947794,7667	37	1	3	24	13
38Q	5733	51	0,9931	0,0005	-390,4676	226317,2564	1080838450,6754		226745975,2218	38	1	2	25	51
38V	1201	111	0,9945	0,0014	-731,5210	226630,2837	2379160614,3650		1298322163,6895	38	2	1	25	111
39	6772	593	0,9947	0,0002	-5112,9860	226671,3150	2676733933,6155		297573319,2506	39	1	5	26	593
40	6179	58	0,9958	0,0011	-396,0059	226918,1898	5996844801,7095		3320110868,0940	40	2	1	27	58
41	6121	47	0,9959	0,0001	-443,2871	226936,4538	6416860893,4081		420016091,6986	41	1	14	27	47

La figure 5.43.1 donne une bonne performance du modèle (l'échelle englobe plus de 200 000 cas pour la première question et peu de cas pour les dernières questions). La figure 5.43.2 donnant les probabilités d'arrêt (indépendantes du nombre de données) montre toujours une bonne estimation pour les premières questions et une sous-estimation de la probabilité d'arrêt pour les questions suivantes.

Nous allons maintenant comparer la performance des différents modèles que nous venons de présenter afin de sélectionner le meilleur d'entre eux. Pour déterminer l'apport des différentes variables covariées et choisir entre les modèles, nous utilisons le test du ratio de vraisemblance. Nous comparons par cette méthode la performance des modèles EG avec 1 covarié (type), EG avec 2 covariés (Type et numéro), EG avec 3 covariés (Type, numéro et complexité) et EG avec 4 covariés (Type, numéro, complexité et nbre de questions virales précédentes). La synthèse de résultats est présentée dans le tableau 5.45 ci-dessous.

Tableau 5.45: Comparaison des modèles

			Comparaison du modèle vs modèle ligne précédente	
	LL	Nbre de paramètres	LR	Khi 2 (0,05;1)
EG + 1 covarié (type)	-592815,71	3		
EG+ 2 covariés (type, numéro)	-589968,11	4	5695,20	3,841
EG + 3covariés (type, numéro, complexité)	-588767,03	5	2402,16	3,841
EG + 4 covariés (type, numéro, complexité, nb de questions virales avant)	-588766,79	6	0,48	3,841

Dans ce tableau 5.45, chaque modèle est comparé au modèle de la ligne précédente et diffère de celui de la ligne précédente par un paramètre supplémentaire (correspondant au covarié de plus). Ceci explique le fait que la valeur du Khi 2 est la même pour tous. On constate sur ce tableau que LR > $\chi^2_{0.05,1}$ (soit LR>3,84) pour les modèles avec 2 et 3 covariés, ce qui signifie que le modèle avec 3 covariés est meilleur que le modèle avec 1 et 2 covariés. Cependant, lorsqu'on ajoute la 4ème variable covariée qui est le nombre de questions virales avant, on ne gagne pas en précision (LR = 0,48 est inférieur à 3,841).

En utilisant par conséquent le test du ratio de vraisemblance du tableau 5.45, on peut conclure que le meilleur des modèles testés est le modèle Exponentiel Gamma avec 3 covariés (type, numéro, complexité de la question). Ceci permet d'affirmer que les variables expliquant la probabilité de non réponse à une question sont le type, le numéro et la complexité de la question, le nombre de questions virales avant n'apportant pas de précision supplémentaire. .

Les paramètres du modèle EG+3 covariés retenu sont les suivants :

Tableau 5.46: paramètres du modèle EG + 3 covariés

r	0,2816
alpha	25,3356
B_no	0,4690
B_type	1,6907
B_complexité	-0,0647

r et alpha sont les paramètres du modèle Gamma (introduisant l'hétérogénéité des individus). B_no, B_type et B_complexité étant respectivement les coefficients des variables covariées numéro de la question, type et complexité. L'interprétation de ces paramètres est la suivante :

Plus on avance dans le questionnaire (le numéro de la question augmente dans ce cas), plus la probabilité cumulée d'arrêt à la question augmente. Lorsqu'on a une question virale, la probabilité d'arrêt augmente aussi (versus une question de qualification). En revanche, pour la complexité de la question, plus le nombre de choix de réponses est important, plus la probabilité cumulée de non réponse à la question est faible.

Les écarts entre les données observées et les estimations nous ont poussé à essayer de comprendre ces écarts et à améliorer encore ce modèle.

L'une des caractéristiques de la distribution exponentielle est d'être « sans mémoire », ce qui signifie que la probabilité qu'un événement arrive dans l'intervalle (t, t+Δ t), étant donné qu'il n'est pas arrivé en t, est indépendant du temps ($P(t<T\leq t+\Delta t \mid T>t)=1-e^{-\lambda\Delta t}$)

Or, on peut se demander dans notre cas si cette hypothèse est correcte puisqu'elle signifie que la probabilité de non réponses à une question ne dépend pas de t.

Pour rendre cette probabilité dépendante de t, on utilise la fonction de taux de hasard qui représente le taux instantané de « défaillance » en t conditionnel à la « survie » en t.

$$h(t)=\lim_{\Delta t\to 0} P(t<T\leq t+\Delta t \mid T>t)/\Delta t=\frac{f(t)}{1-F(t)}$$

La probabilité de défaillance dans le prochain intervalle de temps, étant donné la survie en t, est donné par : $P(t<T\leq t+\Delta t \mid T>t)\approx h(t)x\Delta t$

Une généralisation de la distribution exponentielle qui représente ces taux de hasard croissant ou décroissant est la distribution de Weibull, caractérisée par :

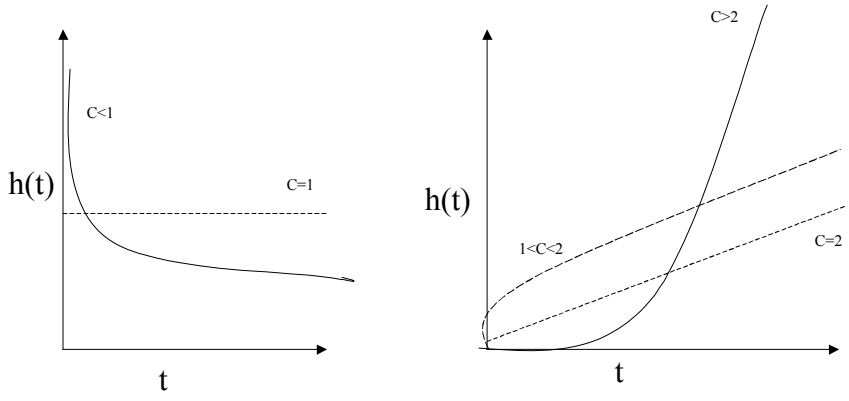
$$F(t)=1-e^{-\lambda t^{c}} \quad \text{avec } \lambda,\ c>0$$

$$h(t)=c\lambda t^{c-1}$$

c est le paramètre de forme et λ le paramètre d'échelle du modèle (dans le cas où c=1, on retrouve le modèle exponentiel).

La fonction de taux de hasard de Weibull peut donc prendre les formes suivantes :

Figure 5.44: la fonction de taux de hasard Weibull



Le taux de hasard est décroissant lorsque c<1 (on a alors une dépendance de durée négative).
Le taux de hasard est croissant lorsque c>1 (on a alors une dépendance de durée positive)

Modèles Weibull-Gamma avec covariés :

En appliquant ce modèle à nos données, nous avons testé un modèle Weibull Gamma avec covariés, afin de les comparer au modèle exponentiel gamma avec covariés que nous avons trouvé auparavant et déterminer si le modèle Weibull fournit de meilleurs résultats.
Le modèle Weibull Gamma sans covariés s'exprime par la formule suivante (Hardie et Fader, 2001) :

$$P(T \leq t) = 1 - \left(\frac{\alpha}{\alpha + t^c} \right)^r$$

Quand c=1, on retrouve un modèle exponentiel gamma.
Le modèle Weibull Gamma avec covariés que nous souhaitons tester (puisque nous avons déjà montré que le type de question, le numéro de la question ou sa complexité influaient sur la probabilité de non réponse), s'exprime par la formule suivante (Hardie et Fader, 2001) :

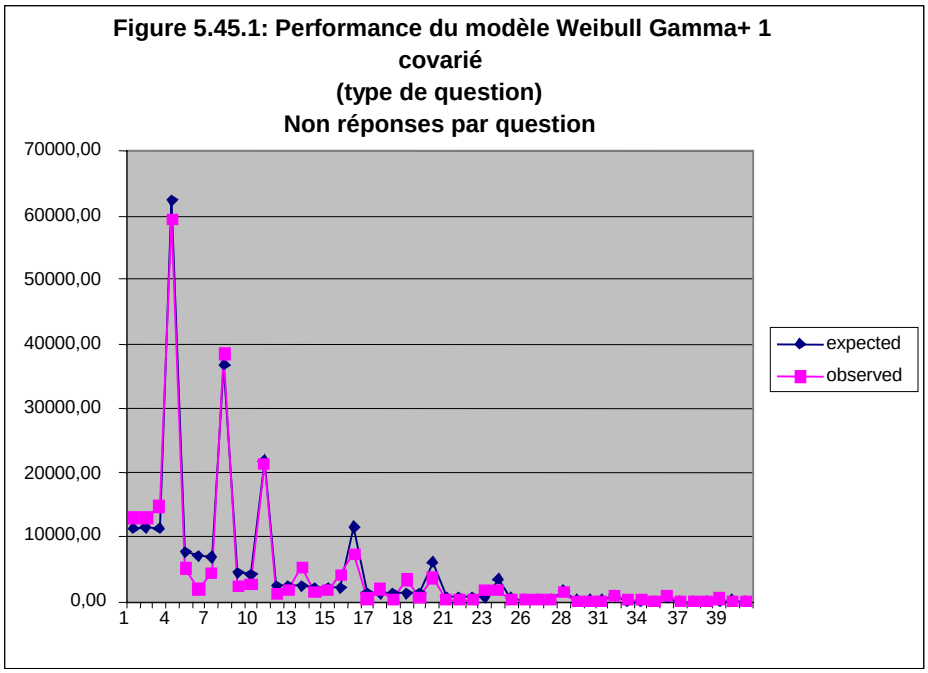
Supposons que le paramètre λ soit distribué sur la population selon une distribution Gamma (pour tenir compte de l'hétérogénéité des individus),

$$P(T \le t) = \int_0^{\infty} (1 - \exp(-\lambda B(t))) \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)} d\lambda = 1 - \left(\frac{\alpha}{\alpha + B(t)} \right)^r$$

avec $B(t) = \sum_{i=1}^{t} \left[i^c - (i-1)^c \right] \exp(\beta' x(i))$

x(i) étant les variables covariées incluses dans le modèle.

Nous avons analysé tout d'abord le modèle WG + 1 covarié : le type de question (qualification ou virale) afin de le comparer au modèle EG + 1 covarié (type question). La performance du modèle est indiqué dans les figures 5.45.1 et 5.45.2, les formules et résultats sur Excel sont fournis dans les tableaux 5.47 et 5.48.



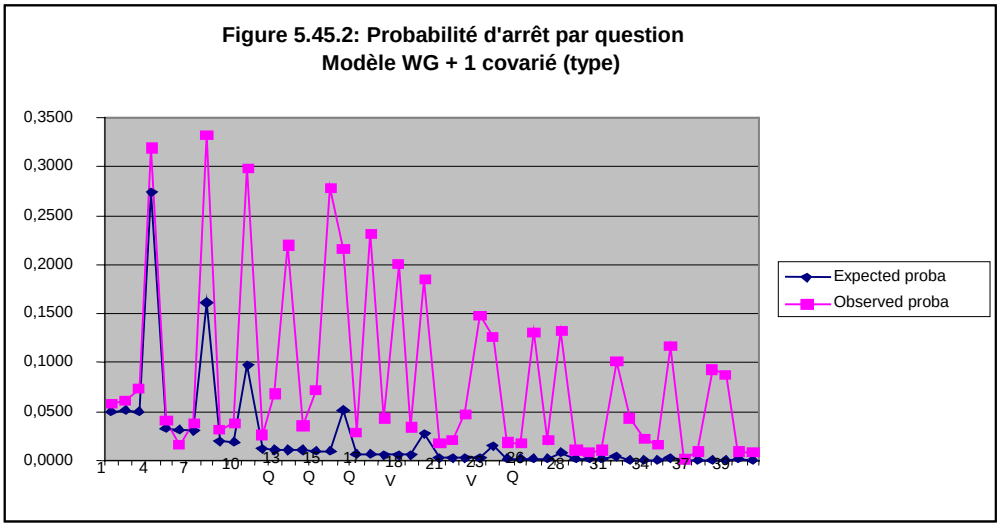


Tableau 5.47 : Modèle Weibull Gamma + 1 covarié (type) : Formules

nbre tot 227879

r	325,260101292924
alpha	42968,8517148798
c	1,06451355274718
B_type	1,89778462259576

LL= =SOMME(F9:F56) =TRANSPOSE(F4)

Questio	Nbre envoi	Nbre non réponses	P(T<=t)	P(non r	LL	E(T(t))	B(t)	exp(BX)	type	PAR QUES Observed
1	227879	13111	=1-(F$2/(F$2+J9))^F$1	=D9	=C9*LN(E9)	=B$1*D9	=A9^F3*K9	=EXP(SOMMEPROD(J$6;M9))	1	=C9
2	214768	13146	=1-(F$2/(F$2+J10))^F$1	=D10-D	=C10*LN(E10)	=B$1*D10	=J9+(A10^F$3-A9^F$3)*K10	=EXP(SOMMEPROD(J$6;M10))	1	=C10
3	201622	14729	=1-(F$2/(F$2+J11))^F$1	=D11-D	=C11*LN(E11)	=B$1*D11	=J10+(A11^F$3-A10^F$3)*K11	=EXP(SOMMEPROD(J$6;M11))	1	=C11
4	186893	59518	=1-(F$2/(F$2+J12))^F$1	=D12-D	=C12*LN(E12)	=B$1*D12	=J11+(A12^F$3-A11^F$3)*K12	=EXP(SOMMEPROD(J$6;M12))	2	=C12
5	127375	5216	=1-(F$2/(F$2+J13))^F$1	=D13-D	=C13*LN(E13)	=B$1*D13	=J12+(A13^F$3-A12^F$3)*K13	=EXP(SOMMEPROD(J$6;M13))	1	=C13
6	122159	1896	=1-(F$2/(F$2+J14))^F$1	=D14-D	=C14*LN(E14)	=B$1*D14	=J13+(A14^F$3-A13^F$3)*K14	=EXP(SOMMEPROD(J$6;M14))	1	=C14
7	120263	4546	=1-(F$2/(F$2+J15))^F$1	=D15-D	=C15*LN(E15)	=B$1*D15	=J14+(A15^F$3-A14^F$3)*K15	=EXP(SOMMEPROD(J$6;M15))	1	=C15
8	115717	38349	=1-(F$2/(F$2+J16))^F$1	=D16-D	=C16*LN(E16)	=B$1*D16	=J15+(A16^F$3-A15^F$3)*K16	=EXP(SOMMEPROD(J$6;M16))	2	=C16
9	77368	2431	=1-(F$2/(F$2+J17))^F$1	=D17-D	=C17*LN(E17)	=B$1*D17	=J16+(A17^F$3-A16^F$3)*K17	=EXP(SOMMEPROD(J$6;M17))	1	=C17
10	74937	2810	=1-(F$2/(F$2+J18))^F$1	=D18-D	=C18*LN(E18)	=B$1*D18	=J17+(A18^F$3-A17^F$3)*K18	=EXP(SOMMEPROD(J$6;M18))	1	=C18
11	72127	21460	=1-(F$2/(F$2+J19))^F$1	=D19-D	=C19*LN(E19)	=B$1*D19	=J18+(A19^F$3-A18^F$3)*K19	=EXP(SOMMEPROD(J$6;M19))	2	=C19
12	50667	1305	=1-(F$2/(F$2+J20))^F$1	=D20-D	=C20*LN(E20)	=B$1*D20	=J19+(A20^F$3-A19^F$3)*K20	=EXP(SOMMEPROD(J$6;M20))	1	=C20
13	24551	1679	=1-(F$2/(F$2+J21))^F$1	=D21-D	=C21*LN(E21)	=B$1*D21	=J20+(A21^F$3-A20^F$3)*K21	=EXP(SOMMEPROD(J$6;M21))	1	=C21
13	24811	5457	=1-(F$2/(F$2+J22))^F$1	=D22-D	=C22*LN(E22)	=B$1*D22	=J21+(A22^F$3-A21^F$3)*K22	=EXP(SOMMEPROD(J$6;M22))	2	=C22
14	42226	1498	=1-(F$2/(F$2+J23))^F$1	=D23-D	=C23*LN(E23)	=B$1*D23	=J22+(A23^F$3-A22^F$3)*K23	=EXP(SOMMEPROD(J$6;M23))	1	=C23
15	25501	1833	=1-(F$2/(F$2+J24))^F$1	=D24-D	=C24*LN(E24)	=B$1*D24	=J23+(A24^F$3-A23^F$3)*K24	=EXP(SOMMEPROD(J$6;M24))	1	=C24
15	15227	4228	=1-(F$2/(F$2+J25))^F$1	=D25-D	=C25*LN(E25)	=B$1*D25	=J24+(A25^F$3-A24^F$3)*K25	=EXP(SOMMEPROD(J$6;M25))	2	=C25
16	34667	7488	=1-(F$2/(F$2+J26))^F$1	=D26-D	=C26*LN(E26)	=B$1*D26	=J25+(A26^F$3-A25^F$3)*K26	=EXP(SOMMEPROD(J$6;M26))	2	=C26
17	18325	531	=1-(F$2/(F$2+J27))^F$1	=D27-D	=C27*LN(E27)	=B$1*D27	=J26+(A27^F$3-A26^F$3)*K27	=EXP(SOMMEPROD(J$6;M27))	1	=C27
17	8854	2048	=1-(F$2/(F$2+J28))^F$1	=D28-D	=C28*LN(E28)	=B$1*D28	=J27+(A28^F$3-A27^F$3)*K28	=EXP(SOMMEPROD(J$6;M28))	2	=C28
18	7273	314	=1-(F$2/(F$2+J29))^F$1	=D29-D	=C29*LN(E29)	=B$1*D29	=J28+(A29^F$3-A28^F$3)*K29	=EXP(SOMMEPROD(J$6;M29))	1	=C29
18	17327	3462	=1-(F$2/(F$2+J30))^F$1	=D30-D	=C30*LN(E30)	=B$1*D30	=J29+(A30^F$3-A29^F$3)*K30	=EXP(SOMMEPROD(J$6;M30))	2	=C30
19	20824	719	=1-(F$2/(F$2+J31))^F$1	=D31-D	=C31*LN(E31)	=B$1*D31	=J30+(A31^F$3-A30^F$3)*K31	=EXP(SOMMEPROD(J$6;M31))	1	=C31
20	20105	3692	=1-(F$2/(F$2+J32))^F$1	=D32-D	=C32*LN(E32)	=B$1*D32	=J31+(A32^F$3-A31^F$3)*K32	=EXP(SOMMEPROD(J$6;M32))	2	=C32
21	16413	287	=1-(F$2/(F$2+J33))^F$1	=D33-D	=C33*LN(E33)	=B$1*D33	=J32+(A33^F$3-A32^F$3)*K33	=EXP(SOMMEPROD(J$6;M33))	1	=C33
22	16126	329	=1-(F$2/(F$2+J34))^F$1	=D34-D	=C34*LN(E34)	=B$1*D34	=J33+(A34^F$3-A33^F$3)*K34	=EXP(SOMMEPROD(J$6;M34))	1	=C34
23	4113	196	=1-(F$2/(F$2+J35))^F$1	=D35-D	=C35*LN(E35)	=B$1*D35	=J34+(A35^F$3-A34^F$3)*K35	=EXP(SOMMEPROD(J$6;M35))	1	=C35
23	11624	1710	=1-(F$2/(F$2+J36))^F$1	=D36-D	=C36*LN(E36)	=B$1*D36	=J35+(A36^F$3-A35^F$3)*K36	=EXP(SOMMEPROD(J$6;M36))	2	=C36
24	13831	1738	=1-(F$2/(F$2+J37))^F$1	=D37-D	=C37*LN(E37)	=B$1*D37	=J36+(A37^F$3-A36^F$3)*K37	=EXP(SOMMEPROD(J$6;M37))	2	=C37
25	12093	214	=1-(F$2/(F$2+J38))^F$1	=D38-D	=C38*LN(E38)	=B$1*D38	=J37+(A38^F$3-A37^F$3)*K38	=EXP(SOMMEPROD(J$6;M38))	1	=C38
26	10130	177	=1-(F$2/(F$2+J39))^F$1	=D39-D	=C39*LN(E39)	=B$1*D39	=J38+(A39^F$3-A38^F$3)*K39	=EXP(SOMMEPROD(J$6;M39))	1	=C39
26	1749	229	=1-(F$2/(F$2+J40))^F$1	=D40-D	=C40*LN(E40)	=B$1*D40	=J39+(A40^F$3-A39^F$3)*K40	=EXP(SOMMEPROD(J$6;M40))	2	=C40
27	11473	247	=1-(F$2/(F$2+J41))^F$1	=D41-D	=C41*LN(E41)	=B$1*D41	=J40+(A41^F$3-A40^F$3)*K41	=EXP(SOMMEPROD(J$6;M41))	1	=C41
28	11226	1475	=1-(F$2/(F$2+J42))^F$1	=D42-D	=C42*LN(E42)	=B$1*D42	=J41+(A42^F$3-A41^F$3)*K42	=EXP(SOMMEPROD(J$6;M42))	2	=C42
29	9751	99	=1-(F$2/(F$2+J43))^F$1	=D43-D	=C43*LN(E43)	=B$1*D43	=J42+(A43^F$3-A42^F$3)*K43	=EXP(SOMMEPROD(J$6;M43))	1	=C43
30	9652	80	=1-(F$2/(F$2+J44))^F$1	=D44-D	=C44*LN(E44)	=B$1*D44	=J43+(A44^F$3-A43^F$3)*K44	=EXP(SOMMEPROD(J$6;M44))	1	=C44
31	9572	101	=1-(F$2/(F$2+J45))^F$1	=D45-D	=C45*LN(E45)	=B$1*D45	=J44+(A45^F$3-A44^F$3)*K45	=EXP(SOMMEPROD(J$6;M45))	1	=C45
32	9471	951	=1-(F$2/(F$2+J46))^F$1	=D46-D	=C46*LN(E46)	=B$1*D46	=J45+(A46^F$3-A45^F$3)*K46	=EXP(SOMMEPROD(J$6;M46))	2	=C46
33	8520	368	=1-(F$2/(F$2+J47))^F$1	=D47-D	=C47*LN(E47)	=B$1*D47	=J46+(A47^F$3-A46^F$3)*K47	=EXP(SOMMEPROD(J$6;M47))	1	=C47
34	8152	177	=1-(F$2/(F$2+J48))^F$1	=D48-D	=C48*LN(E48)	=B$1*D48	=J47+(A48^F$3-A47^F$3)*K48	=EXP(SOMMEPROD(J$6;M48))	1	=C48
35	7975	120	=1-(F$2/(F$2+J49))^F$1	=D49-D	=C49*LN(E49)	=B$1*D49	=J48+(A49^F$3-A48^F$3)*K49	=EXP(SOMMEPROD(J$6;M49))	1	=C49
36	7855	908	=1-(F$2/(F$2+J50))^F$1	=D50-D	=C50*LN(E50)	=B$1*D50	=J49+(A50^F$3-A49^F$3)*K50	=EXP(SOMMEPROD(J$6;M50))	2	=C50
37	6947	13	=1-(F$2/(F$2+J51))^F$1	=D51-D	=C51*LN(E51)	=B$1*D51	=J50+(A51^F$3-A50^F$3)*K51	=EXP(SOMMEPROD(J$6;M51))	1	=C51
38	5733	51	=1-(F$2/(F$2+J52))^F$1	=D52-D	=C52*LN(E52)	=B$1*D52	=J51+(A52^F$3-A51^F$3)*K52	=EXP(SOMMEPROD(J$6;M52))	1	=C52
38	1201	111	=1-(F$2/(F$2+J53))^F$1	=D53-D	=C53*LN(E53)	=B$1*D53	=J52+(A53^F$3-A52^F$3)*K53	=EXP(SOMMEPROD(J$6;M53))	2	=C53
39	6772	593	=1-(F$2/(F$2+J54))^F$1	=D54-D	=C54*LN(E54)	=B$1*D54	=J53+(A54^F$3-A53^F$3)*K54	=EXP(SOMMEPROD(J$6;M54))	1	=C54
40	6179	58	=1-(F$2/(F$2+J55))^F$1	=D55-D	=C55*LN(E55)	=B$1*D55	=J54+(A55^F$3-A54^F$3)*K55	=EXP(SOMMEPROD(J$6;M55))	2	=C55
41	6121	47	=1-(F$2/(F$2+J56))^F$1	=D56-D	=C56*LN(E56)	=B$1*D56	=J55+(A56^F$3-A55^F$3)*K56	=EXP(SOMMEPROD(J$6;M56))	1	=C56

Tableau 5.48 : Modèle Weibull Gamma + 1 covarié (type) : Résultats

nbre total de	227879			r	325,2601						
				alpha	42968,8517						
				c	1,0645						
				B_type	1,8978						
				LL=	-586781,3882			1,8978			
Questions	Nbre envois	Nbre non réponses	P(T<=t)	P(non reponse Ques	LL	E(T(t))	B(t)	exp(BX)	Covar type	PAR QUESTION Observed	
1	227879	13111	0,0492	0,0492	-39477,7376	11220,86942	6,6711	6,6711	1	13111,00	
2	214768	13146	0,1002	0,0510	-39128,5414	22836,53791	13,9524	6,6711	1	13146,00	
3	201622	14729	0,1501	0,0498	-44171,7245	34193,73772	21,4832	6,6711	1	14729,00	
4	186893	59518	0,4236	0,2735	-77162,2820	96518,69315	72,8351	44,5036	2	59518,00	
5	127375	5216	0,4566	0,0331	-17778,5491	104059,6616	80,6592	6,6711	1	5216,00	
6	122159	1896	0,4882	0,0316	-6550,9785	111256,6119	88,5856	6,6711	1	1896,00	
7	120263	4546	0,5183	0,0301	-15932,5488	118105,4199	96,5981	6,6711	1	4546,00	
8	115717	38349	0,6794	0,1611	-70007,7363	154823,6037	150,5466	44,5036	2	38349,00	
9	77368	2431	0,6985	0,0191	-9619,5917	159180,5155	158,6992	6,6711	1	2431,00	
10	74937	2810	0,7166	0,0181	-11273,0537	163305,4573	166,9105	6,6711	1	2810,00	
11	72127	21460	0,8130	0,0963	-50214,7080	185258,2339	222,0441	44,5036	2	21460,00	
12	50667	1305	0,8243	0,0113	-5844,5957	187844,454	230,3573	6,6711	1	1305,00	
13	24551	1679	0,8350	0,0107	-7616,2600	190285,9882	238,7154	6,6711	1	1679,00	
13	24811	5457	0,8350	0,0107	-24753,9790	190285,9882	238,7154	44,5036	2	5457,00	
14	42226	1498	0,8451	0,0101	-6882,5435	192589,2504	247,1152	6,6711	1	1498,00	
15	25501	1833	0,8547	0,0095	-8529,7469	194760,6664	255,5537	6,6711	1	1833,00	
15	15227	4228	0,8547	0,0095	-19674,7245	194760,6664	255,5537	44,5036	2	4228,00	
16	34667	7488	0,9050	0,0503	-22382,2169	206230,6865	312,0909	44,5036	2	7488,00	
17	18325	531	0,9109	0,0059	-2726,9400	207571,5778	320,6001	6,6711	1	531,00	
17	8854	2048	0,9109	0,0059	-10517,4634	207571,5778	320,6001	44,5036	2	2048,00	
18	7273	314	0,9164	0,0055	-1631,5237	208833,8072	329,1416	6,6711	1	314,00	
18	17327	3462	0,9164	0,0055	-17988,3279	208833,8072	329,1416	44,5036	2	3462,00	
19	20824	719	0,9216	0,0052	-3779,6582	210021,4711	337,7139	6,6711	1	719,00	
20	20105	3692	0,9491	0,0274	-13277,9017	216270,4415	395,0948	44,5036	2	3692,00	
21	16413	287	0,9523	0,0032	-1649,3776	216997,9386	403,7241	6,6711	1	287,00	
22	16126	329	0,9553	0,0030	-1911,1272	217681,7437	412,3799	6,6711	1	329,00	
23	4113	196	0,9581	0,0028	-1150,7467	218324,2737	421,0611	6,6711	1	196,00	
23	11624	1710	0,9581	0,0028	-10039,6776	218324,2737	421,0611	44,5036	2	1710,00	
24	13831	1738	0,9729	0,0148	-7323,4248	221694,9271	479,1370	44,5036	2	1738,00	
25	12093	214	0,9746	0,0017	-1362,6384	222086,0809	487,8661	6,6711	1	214,00	
26	10130	177	0,9762	0,0016	-1138,1997	222453,3384	496,6177	6,6711	1	177,00	
26	1749	229	0,9762	0,0016	-1472,5861	222453,3384	496,6177	44,5036	2	229,00	
27	11473	247	0,9777	0,0015	-1603,9679	222798,0724	505,3910	6,6711	1	247,00	
28	11226	1475	0,9856	0,0079	-7137,1055	224602,2398	564,0590	44,5036	2	1475,00	
29	9751	99	0,9865	0,0009	-692,5090	224811,0717	572,8736	6,6711	1	99,00	
30	9652	80	0,9874	0,0009	-564,7149	225006,9775	581,7079	6,6711	1	80,00	
31	9572	101	0,9882	0,0008	-719,4269	225190,7193	590,5612	6,6711	1	101,00	
32	9471	951	0,9924	0,0042	-5201,8698	226150,4719	649,7455	44,5036	2	951,00	
33	8520	368	0,9929	0,0005	-2807,2258	226261,3286	658,6351	6,6711	1	368,00	
34	8152	177	0,9934	0,0005	-1361,6468	226365,2516	667,5422	6,6711	1	177,00	
35	7975	120	0,9938	0,0004	-930,9217	226462,6578	676,4661	6,6711	1	120,00	
36	7855	908	0,9960	0,0022	-5544,3909	226970,6228	736,1086	44,5036	2	908,00	
37	6947	13	0,9963	0,0003	-107,4628	227029,1917	745,0651	6,6711	1	13,00	
38	5733	51	0,9965	0,0002	-424,9080	227084,0658	754,0372	6,6711	1	51,00	
38	1201	111	0,9965	0,0002	-924,7997	227084,0658	754,0372	44,5036	2	111,00	
39	6772	593	0,9967	0,0002	-4979,3242	227135,4706	763,0245	6,6711	1	593,00	
40	6179	58	0,9979	0,0012	-391,3080	227403,1743	823,0793	44,5036	2	58,00	
41	6121	47	0,9980	0,0001	-418,6953	227433,9942	832,0961	6,6711	1	47,00	

Afin de déterminer si le modèle WG+1covarié est meilleur que le modèle EG+1 covarié, nous testons donc si c=1 (les 2 modèles ne sont pas différents) par le test du ratio de vraisemblance :

Tableau 5.49: Comparaison des modèles

		nbr paramètres
LL Exponentiel Gamma + 1 Covariés (type)	-592815,71	3
LL Weibull Gamma + 1 Covariés (type)	-586781,39	4
LR(df=1)	12068,64	
Chi-2 (0,95,1)	3,84	

Le tableau 5.49 montre que le modèle WG + 1 covarié est meilleur que le modèle EG + 1 covarié, puisque LR> 3,84. Le modèle de Weibull permet d'améliorer la performance de l'estimation des données par rapport aux données observées.

Nous avons ensuite effectué la même démarche pour tester le modèle WG+3 covariés (type, numéro, complexité de la question) et le comparer au modèle Exponentiel Gamma avec les 3 variables covariées. La comparaison des modèles obtenus sous Excel par la méthode du maximum de vraisemblance donne les résultats suivants :

Tableau 5.50: Comparaison des modèles

		nbr paramètres
LL Exponentiel Gamma + 3 Covariés (type, numéro, complexité)	-587411,43	5
LL Weibull Gamma + 3 Covariés (type, numéro, complexité)	-585006,72	6
LR(df=1)	4809,44	
Chi-2 (0,95,1)	3,84	

On en conclut que le modèle Weibull Gamma avec 3 covariés est plus performant que le modèle Exponentiel Gamma avec 3 covariés (type, numéro et complexité de la question), puisque LR > 3,84. Ceci signifie qu'on améliore la précision du modèle en optant pour le modèle Weibull Gamma avec 3 covariés plutôt qu'Exponentiel Gamma avec les mêmes 3 covariés.

Afin d'être exhaustif dans notre démarche, nous avons testé sur le modèle Weibull Gamma la 4ème variable covariée qui est le nombre de questions virales précédemment répondues.

Tableau 5.51 : Modèle Weibull Gamma + 4 covariés: Formules

227879

Tableau 5.5

r	0,499900503052051
alpha	38,0916004397928
c	0,668132232474362
B_no	1,00000000000057
B_type	1,71137084823235
B_qvira	0,513824386860274
B_complex	0,0000001

LL= =SOMME(F12:F59) =T =TRANSPOSE(F6) =TRANSPOSE(F7) =TRA

Nbre cas	Nbre d	P(T<=t)	P(non repor	LL	E(T(t))	B(t)	exp(BX)	Cova no	type	Nbre	com	PAR QUI Observec
227879	13111	=1-(F$3/(F$3+J12))^F$2	=D12	=C12*LN(E12)	=B$2*D12	=A12^F4*K12	=EXP(SOMMEPROD(I$9:L$9;L12:O12))	1	1	0	5	=C12
214768	13146	=1-(F$3/(F$3+J13))^F$2	=D13-D12	=C13*LN(E13)	=B$2*D13	=J12+(A13^F$4-A12^F$4)*K13	=EXP(SOMMEPROD(I$9:L$9;L13:O13))	2	1	1	11	=C13
201622	14729	=1-(F$3/(F$3+J14))^F$2	=D14-D13	=C14*LN(E14)	=B$2*D14	=J13+(A14^F$4-A13^F$4)*K14	=EXP(SOMMEPROD(I$9:L$9;L14:O14))	3	1	2	9	=C14
186893	59518	=1-(F$3/(F$3+J15))^F$2	=D15-D14	=C15*LN(E15)	=B$2*D15	=J14+(A15^F$4-A14^F$4)*K15	=EXP(SOMMEPROD(I$9:L$9;L15:O15))	4	2	3	1	=C15
127375	5216	=1-(F$3/(F$3+J16))^F$2	=D16-D15	=C16*LN(E16)	=B$2*D16	=J15+(A16^F$4-A15^F$4)*K16	=EXP(SOMMEPROD(I$9:L$9;L16:O16))	5	1	3	6	=C16
122159	1896	=1-(F$3/(F$3+J17))^F$2	=D17-D16	=C17*LN(E17)	=B$2*D17	=J16+(A17^F$4-A16^F$4)*K17	=EXP(SOMMEPROD(I$9:L$9;L17:O17))	6	1	4	15	=C17
120263	4546	=1-(F$3/(F$3+J18))^F$2	=D18-D17	=C18*LN(E18)	=B$2*D18	=J17+(A18^F$4-A17^F$4)*K18	=EXP(SOMMEPROD(I$9:L$9;L18:O18))	7	1	5	13	=C18
115717	38349	=1-(F$3/(F$3+J19))^F$2	=D19-D18	=C19*LN(E19)	=B$2*D19	=J18+(A19^F$4-A18^F$4)*K19	=EXP(SOMMEPROD(I$9:L$9;L19:O19))	8	2	6	1	=C19
77368	2431	=1-(F$3/(F$3+J20))^F$2	=D20-D19	=C20*LN(E20)	=B$2*D20	=J19+(A20^F$4-A19^F$4)*K20	=EXP(SOMMEPROD(I$9:L$9;L20:O20))	9	1	6	5	=C20
74937	2810	=1-(F$3/(F$3+J21))^F$2	=D21-D20	=C21*LN(E21)	=B$2*D21	=J20+(A21^F$4-A20^F$4)*K21	=EXP(SOMMEPROD(I$9:L$9;L21:O21))	10	1	7	7	=C21
72127	21460	=1-(F$3/(F$3+J22))^F$2	=D22-D21	=C22*LN(E22)	=B$2*D22	=J21+(A22^F$4-A21^F$4)*K22	=EXP(SOMMEPROD(I$9:L$9;L22:O22))	11	2	8	1	=C22
50667	1305	=1-(F$3/(F$3+J23))^F$2	=D23-D22	=C23*LN(E23)	=B$2*D23	=J22+(A23^F$4-A22^F$4)*K23	=EXP(SOMMEPROD(I$9:L$9;L23:O23))	12	1	8	9	=C23
24551	1679	=1-(F$3/(F$3+J24))^F$2	=D24-D23	=C24*LN(E24)	=B$2*D24	=J23+(A24^F$4-A23^F$4)*K24	=EXP(SOMMEPROD(I$9:L$9;L24:O24))	13	1	9,115	9	=C24
24811	5457	=1-(F$3/(F$3+J25))^F$2	=D25-D23	=C25*LN(E25)	=B$2*D25	=J24+(A25^F$4-A24^F$4)*K25	=EXP(SOMMEPROD(I$9:L$9;L25:O25))	13	2	9	1	=C25
42226	1498	=1-(F$3/(F$3+J26))^F$2	=D26-D25	=C26*LN(E26)	=B$2*D26	=J25+(A26^F$4-A25^F$4)*K26	=EXP(SOMMEPROD(I$9:L$9;L26:O26))	14	1	9,555	6	=C26
25501	1833	=1-(F$3/(F$3+J27))^F$2	=D27-D26	=C27*LN(E27)	=B$2*D27	=J26+(A27^F$4-A26^F$4)*K27	=EXP(SOMMEPROD(I$9:L$9;L27:O27))	15	1	10,37	13	=C27
15227	4228	=1-(F$3/(F$3+J28))^F$2	=D28-D26	=C28*LN(E28)	=B$2*D28	=J27+(A28^F$4-A27^F$4)*K28	=EXP(SOMMEPROD(I$9:L$9;L28:O28))	15	2	10,84	1	=C28
34667	7488	=1-(F$3/(F$3+J29))^F$2	=D29-D28	=C29*LN(E29)	=B$2*D29	=J28+(A29^F$4-A28^F$4)*K29	=EXP(SOMMEPROD(I$9:L$9;L29:O29))	16	2	11,18	1	=C29
18325	531	=1-(F$3/(F$3+J30))^F$2	=D30-D29	=C30*LN(E30)	=B$2*D30	=J29+(A30^F$4-A29^F$4)*K30	=EXP(SOMMEPROD(I$9:L$9;L30:O30))	17	1	11,31	14	=C30
8854	2048	=1-(F$3/(F$3+J31))^F$2	=D31-D29	=C31*LN(E31)	=B$2*D31	=J30+(A31^F$4-A30^F$4)*K31	=EXP(SOMMEPROD(I$9:L$9;L31:O31))	17	2	10,83	1	=C31
7273	314	=1-(F$3/(F$3+J32))^F$2	=D32-D31	=C32*LN(E32)	=B$2*D32	=J31+(A32^F$4-A31^F$4)*K32	=EXP(SOMMEPROD(I$9:L$9;L32:O32))	18	1	11,88	6	=C32
17327	3462	=1-(F$3/(F$3+J33))^F$2	=D33-D31	=C33*LN(E33)	=B$2*D33	=J32+(A33^F$4-A32^F$4)*K33	=EXP(SOMMEPROD(I$9:L$9;L33:O33))	18	2	11,92	1	=C33
20824	719	=1-(F$3/(F$3+J34))^F$2	=D34-D33	=C34*LN(E34)	=B$2*D34	=J33+(A34^F$4-A33^F$4)*K34	=EXP(SOMMEPROD(I$9:L$9;L34:O34))	19	1	12,23	7	=C34
20105	3692	=1-(F$3/(F$3+J35))^F$2	=D35-D34	=C35*LN(E35)	=B$2*D35	=J34+(A35^F$4-A34^F$4)*K35	=EXP(SOMMEPROD(I$9:L$9;L35:O35))	20	2	13,23	1	=C35
16413	287	=1-(F$3/(F$3+J36))^F$2	=D36-D35	=C36*LN(E36)	=B$2*D36	=J35+(A36^F$4-A35^F$4)*K36	=EXP(SOMMEPROD(I$9:L$9;L36:O36))	21	1	13,21	11	=C36
16126	329	=1-(F$3/(F$3+J37))^F$2	=D37-D36	=C37*LN(E37)	=B$2*D37	=J36+(A37^F$4-A36^F$4)*K37	=EXP(SOMMEPROD(I$9:L$9;L37:O37))	22	1	14,21	4	=C37
4113	196	=1-(F$3/(F$3+J38))^F$2	=D38-D37	=C38*LN(E38)	=B$2*D38	=J37+(A38^F$4-A37^F$4)*K38	=EXP(SOMMEPROD(I$9:L$9;L38:O38))	23	1	15,44	3	=C38
11624	1710	=1-(F$3/(F$3+J39))^F$2	=D39-D37	=C39*LN(E39)	=B$2*D39	=J38+(A39^F$4-A38^F$4)*K39	=EXP(SOMMEPROD(I$9:L$9;L39:O39))	23	2	15,1	1	=C39
13831	1738	=1-(F$3/(F$3+J40))^F$2	=D40-D39	=C40*LN(E40)	=B$2*D40	=J39+(A40^F$4-A39^F$4)*K40	=EXP(SOMMEPROD(I$9:L$9;L40:O40))	24	2	15,5	1	=C40
12093	214	=1-(F$3/(F$3+J41))^F$2	=D41-D40	=C41*LN(E41)	=B$2*D41	=J40+(A41^F$4-A40^F$4)*K41	=EXP(SOMMEPROD(I$9:L$9;L41:O41))	25	1	15,48	3	=C41
10130	177	=1-(F$3/(F$3+J42))^F$2	=D42-D41	=C42*LN(E42)	=B$2*D42	=J41+(A42^F$4-A41^F$4)*K42	=EXP(SOMMEPROD(I$9:L$9;L42:O42))	26	1	15,48	18	=C42
1749	229	=1-(F$3/(F$3+J43))^F$2	=D43-D41	=C43*LN(E43)	=B$2*D43	=J42+(A43^F$4-A42^F$4)*K43	=EXP(SOMMEPROD(I$9:L$9;L43:O43))	26	2	16,51	1	=C43
11473	247	=1-(F$3/(F$3+J44))^F$2	=D44-D43	=C44*LN(E44)	=B$2*D44	=J43+(A44^F$4-A43^F$4)*K44	=EXP(SOMMEPROD(I$9:L$9;L44:O44))	27	1	17,35	15	=C44
11226	1475	=1-(F$3/(F$3+J45))^F$2	=D45-D44	=C45*LN(E45)	=B$2*D45	=J44+(A45^F$4-A44^F$4)*K45	=EXP(SOMMEPROD(I$9:L$9;L45:O45))	28	2	18,36	1	=C45
9751	99	=1-(F$3/(F$3+J46))^F$2	=D46-D45	=C46*LN(E46)	=B$2*D46	=J45+(A46^F$4-A45^F$4)*K46	=EXP(SOMMEPROD(I$9:L$9;L46:O46))	29	1	18,33	15	=C46
9652	80	=1-(F$3/(F$3+J47))^F$2	=D47-D46	=C47*LN(E47)	=B$2*D47	=J46+(A47^F$4-A46^F$4)*K47	=EXP(SOMMEPROD(I$9:L$9;L47:O47))	30	1	19,34	4	=C47
9572	101	=1-(F$3/(F$3+J48))^F$2	=D48-D47	=C48*LN(E48)	=B$2*D48	=J47+(A48^F$4-A47^F$4)*K48	=EXP(SOMMEPROD(I$9:L$9;L48:O48))	31	1	20,34	5	=C48
9471	951	=1-(F$3/(F$3+J49))^F$2	=D49-D48	=C49*LN(E49)	=B$2*D49	=J48+(A49^F$4-A48^F$4)*K49	=EXP(SOMMEPROD(I$9:L$9;L49:O49))	32	2	21,36	1	=C49
8520	368	=1-(F$3/(F$3+J50))^F$2	=D50-D49	=C50*LN(E50)	=B$2*D50	=J49+(A50^F$4-A49^F$4)*K50	=EXP(SOMMEPROD(I$9:L$9;L50:O50))	33	1	21,49	2	=C50
8152	177	=1-(F$3/(F$3+J51))^F$2	=D51-D50	=C51*LN(E51)	=B$2*D51	=J50+(A51^F$4-A50^F$4)*K51	=EXP(SOMMEPROD(I$9:L$9;L51:O51))	34	1	22,08	15	=C51
7975	120	=1-(F$3/(F$3+J52))^F$2	=D52-D51	=C52*LN(E52)	=B$2*D52	=J51+(A52^F$4-A51^F$4)*K52	=EXP(SOMMEPROD(I$9:L$9;L52:O52))	35	1	23,08	3	=C52
7855	908	=1-(F$3/(F$3+J53))^F$2	=D53-D52	=C53*LN(E53)	=B$2*D53	=J52+(A53^F$4-A52^F$4)*K53	=EXP(SOMMEPROD(I$9:L$9;L53:O53))	36	2	24,11	1	=C53
6947	13	=1-(F$3/(F$3+J54))^F$2	=D54-D53	=C54*LN(E54)	=B$2*D54	=J53+(A54^F$4-A53^F$4)*K54	=EXP(SOMMEPROD(I$9:L$9;L54:O54))	37	1	24,08	3	=C54
5733	51	=1-(F$3/(F$3+J55))^F$2	=D55-D54	=C55*LN(E55)	=B$2*D55	=J54+(A55^F$4-A54^F$4)*K55	=EXP(SOMMEPROD(I$9:L$9;L55:O55))	38	1	25,15	2	=C55
1201	111	=1-(F$3/(F$3+J56))^F$2	=D56-D54	=C56*LN(E56)	=B$2*D56	=J55+(A56^F$4-A55^F$4)*K56	=EXP(SOMMEPROD(I$9:L$9;L56:O56))	38	2	24,7	1	=C56
6772	593	=1-(F$3/(F$3+J57))^F$2	=D57-D56	=C57*LN(E57)	=B$2*D57	=J56+(A57^F$4-A56^F$4)*K57	=EXP(SOMMEPROD(I$9:L$9;L57:O57))	39	1	25,92	5	=C57
6179	58	=1-(F$3/(F$3+J58))^F$2	=D58-D57	=C58*LN(E58)	=B$2*D58	=J57+(A58^F$4-A57^F$4)*K58	=EXP(SOMMEPROD(I$9:L$9;L58:O58))	40	2	26,94	1	=C58
6121	47	=1-(F$3/(F$3+J59))^F$2	=D59-D58	=C59*LN(E59)	=B$2*D59	=J58+(A59^F$4-A58^F$4)*K59	=EXP(SOMMEPROD(I$9:L$9;L59:O59))	41		26,94	14	=C59

Tableau 5.52: Modèle Weibull Gamma + 4 covariés: résultats

nbre 227879

r	0,4999
alpha	38,0916
c	0,6681
B_no	0,0000
B_type	1,7114
B_qvira	0,5138
B_comple	0,0000

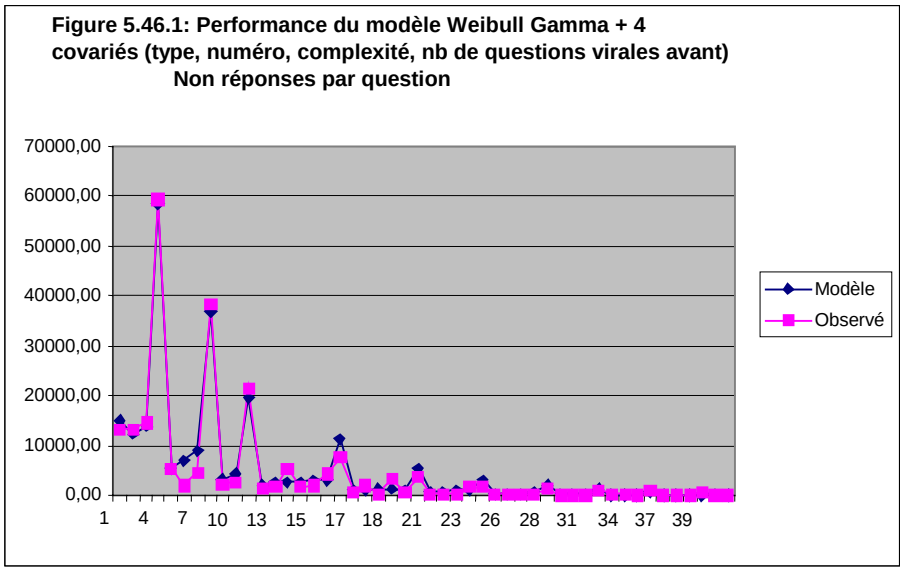
LL= -584005,7037

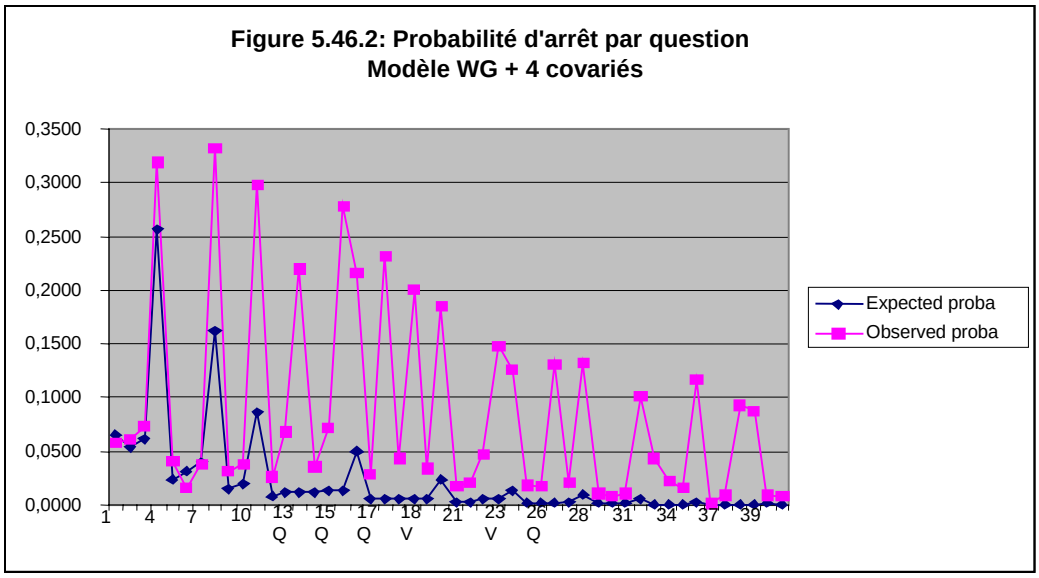
1E-07 1,711371 0,513824 1E-07

Covariates

Ques	Nbre cas	Nbre de n	P(T<=t)	P(non rep	LL	E(T(t))	B(t)	exp(BX)	no	type	Nbre q vira	complexité	PAR QUESTION Observed	Expected
1	227879	13111	0,0656	0,0656	-35718,5447	14946,78	5,536549	5,536549	1	1	0,00	5	13111,00	14946,78
2	214768	13146	0,1190	0,0534	-38514,2516	27118,11	10,98806	9,255302	2	1	1,00	11	13146,00	12171,33
3	201622	14729	0,1805	0,0615	-41066,2212	41141,05	18,63766	15,47183	3	1	2,00	9	14729,00	14022,94
4	186893	59518	0,4364	0,2559	-81125,5066	99451,03	81,86323	143,1961	4	2	3,00	1	59518,00	58309,98
5	127375	5216	0,4596	0,0232	-19641,9991	104726,6	92,3628	25,86382	5	1	3,00	6	5216,00	5275,61
6	122159	1896	0,4907	0,0311	-6580,6938	111811,7	108,7789	43,23587	6	1	4,00	15	1896,00	7085,03
7	120263	4546	0,5305	0,0398	-14656,0767	120880,7	134,7367	72,27619	7	1	5,00	13	4546,00	9069,03
8	115717	38349	0,6921	0,1616	-69893,9778	157708	363,8155	668,9364	8	2	6,00	1	38349,00	36827,27
9	77368	2431	0,7062	0,0142	-10349,4113	160935	403,505	120,822	9	1	6,00	5	2431,00	3226,99
10	74937	2810	0,7254	0,0192	-11107,1883	165310,7	467,4451	201,9749	10	1	7,00	7	2810,00	4375,75
11	72127	21460	0,8120	0,0865	-52519,4337	185028	1039,875	1869,335	11	2	8,00	1	21460,00	19717,32
12	50667	1305	0,8201	0,0082	-6271,7518	186892,3	1140,189	337,6359	12	1	8,00	9	1305,00	1864,28
13	24551	1679	0,8321	0,0119	-7438,2296	189606,9	1313,305	599,0359	13	1	9,12	9	1679,00	2714,64
13	24811	5457	0,8321	0,0119	-24175,3538	189606,9	1313,305	3124,916	13	2	9,00	1	5457,00	2714,64
14	42226	1498	0,8438	0,0118	-6653,9346	192289,9	1524,79	750,7484	14	1	9,56	6	1498,00	2683,00
15	25501	1833	0,8575	0,0137	-7866,1004	195408,7	1839,505	1144,03	15	1	10,38	13	1833,00	3118,76
15	15227	4228	0,8575	0,0137	-18143,9567	195408,7	1839,505	8044,407	15	2	10,84	1	4228,00	3118,76
16	34667	7488	0,9075	0,0500	-22434,2625	206799,3	4417,646	9581,726	16	2	11,18	1	7488,00	11390,57
17	18325	531	0,9122	0,0047	-2848,2198	207866,4	4905,584	1851,469	17	1	11,31	14	531,00	1067,09
17	8854	2048	0,9122	0,0047	-10985,2244	207866,4	4905,584	8008,174	17	2	10,83	1	2048,00	1067,09
18	7273	314	0,9174	0,0052	-1651,1232	209052,2	5548,113	2486,155	18	1	11,89	6	314,00	1185,85
18	17327	3462	0,9174	0,0052	-18204,4221	209052,2	5548,113	14028,86	18	2	11,92	1	3462,00	1185,85
19	20824	719	0,9225	0,0051	-3799,2182	210208	6302,913	2974,951	19	1	12,23	7	719,00	1155,79
20	20105	3692	0,9463	0,0238	-13796,5830	215638	13178,1	27575,4	20	2	13,24	1	3692,00	5429,94
21	16413	287	0,9486	0,0023	-1743,9853	216161,2	14384,9	4921,346	21	1	13,21	11	287,00	523,20
22	16126	329	0,9518	0,0032	-1888,7220	216893,1	16371,38	8229,974	22	1	14,22	4	329,00	732,00
23	4113	196	0,9564	0,0046	-1052,8214	217952,1	20059,88	15513,74	23	1	15,45	3	196,00	1058,95
23	11624	1710	0,9564	0,0046	-9185,3299	217952,1	20059,88	71907,09	23	2	15,10	1	1710,00	1058,95
24	13831	1738	0,9694	0,0130	-7546,7620	220916,3	40820,58	88588,72	24	2	15,51	1	1738,00	2964,19
25	12093	214	0,9707	0,0013	-1425,1663	221208,3	44477,31	15821,08	25	1	15,49	3	214,00	292,05
26	10130	177	0,9718	0,0011	-1202,8264	221463,3	48085,8	15821,11	26	1	15,49	18	177,00	254,92
26	1749	229	0,9718	0,0011	-1556,1992	221463,3	48085,8	148226,9	26	2	16,51	1	229,00	254,92
27	11473	247	0,9742	0,0024	-1491,8713	222006	57393,05	41331,1	27	1	17,36	15	247,00	542,73
28	11226	1475	0,9837	0,0094	-6877,1225	224157,9	143050,8	385089,5	28	2	18,37	1	1475,00	2151,92
29	9751	99	0,9845	0,0008	-706,3574	224339,5	158110,8	68512,52	29	1	18,34	15	99,00	181,57
30	9652	80	0,9856	0,0011	-545,3270	224589,1	183026,7	114654,5	30	1	19,34	4	80,00	249,63
31	9572	101	0,9870	0,0014	-664,1216	224906,8	224263,2	191867,9	31	1	20,34	5	101,00	317,70
32	9471	951	0,9921	0,0051	-5017,4542	226071,9	606875,3	1799406	32	2	21,37	1	951,00	1165,14
33	8520	368	0,9925	0,0004	-2846,3874	226171,6	679813,1	346599,8	33	1	21,49	2	368,00	99,67
34	8152	177	0,9930	0,0005	-1350,3837	226282,4	777420,2	468516,7	34	1	22,08	15	177,00	110,75
35	7975	120	0,9936	0,0006	-884,2338	226426,1	938887,4	782650,5	35	1	23,08	3	120,00	143,73
36	7855	908	0,9960	0,0024	-5468,5695	226978,3	2443519	7362620	36	2	24,11	1	908,00	552,20
37	6947	13	0,9962	0,0002	-110,8071	227023,6	2709132	1311760	37	1	24,09	3	13,00	45,28
38	5733	51	0,9965	0,0003	-416,9453	227087,7	3166281	2278035	38	1	25,16	2	51,00	64,15
38	1201	111	0,9965	0,0003	-907,4693	227087,7	3166281	9979999	38	2	24,70	1	111,00	64,15
39	6772	593	0,9968	0,0003	-4774,9311	227160,3	3838022	3376735	39	1	25,93	5	593,00	72,56
40	6179	58	0,9981	0,0012	-389,7660	227435,2	10067677	31583161	40	2	26,95	1	58,00	274,92
41	6121	47	0,9981	0,0000	-510,4592	227439,6	10269150	1029941	41		26,95	14	47,00	4,37

La performance du modèle WG + 4 covariés est la suivante :





Le modèle WG+ 4 covariés donne de bons résultats sur les valeurs brutes de non réponses par question. Lorsqu'on regarde plus précisément les probabilités d'arrêts observées et fournies par le modèle, le modèle ne capte pas bien les probabilités d'arrêt par question, même si par le fonctionnement du jeu tel que nous l'avons décrit précédemment, les probabilités sur les dernières questions portent sur moins d'individus (donc l'erreur du modèle sur ces probabilités est moins « grave » qu'une erreur sur les premières questions).

Tableau 5.53: Comparaison des modèles

		nbr paramètres
LL Exponentiel Gamma + 4 Covariés (type, numéro, complexité, nb de questions virales avant)	-585732,88	6
LL Weibull Gamma + 4 Covariés (type, numéro, complexité, nb de questions virales avant)	-584005,70	7
LR(df=1)	3454,36	
Chi-2 (0,95,1)	3,84	

		nbr paramètres
LL Weibull Gamma + 3 Covariés (type, numéro, complexité)	-585006,72	6
LL Weibull Gamma + 4 Covariés (type, numéro, complexité, nb de questions virales avant)	-584005,70	7
LR(df=1)	2002,03	
Chi-2 (0,95,1)	3,84	

Comme le montre le test du ratio de vraisemblance du tableau 5.53, le modèle WG + 4 covariés est meilleur que le modèle EG + 4 covariés (LR > 3,84). De plus, le modèle WG + 4 covariés est également meilleur que le modèle WG + 3 covariés (LR > 3,84).

Parmi les modèles testés jusque là, on retiendra donc pour l'instant le modèle WG + 4 Covariés pour estimer la probabilité de non réponse pour toute question dont on connaît le type (qualification ou virale), le numéro dans le questionnaire, la complexité (nbre d'items de réponse) et le nbre de questions virales précédentes.

Les paramètres du modèle Weibull Gamma avec 4 covariés sont les suivants :

Tableau 5.54: Paramètres du modèle WG + 4 covariés

r	0,4999005
alpha	38,0916004
c	0,6681322
B_no	0,0000001
B_type	1,7113708
B_qvira	0,5138244
B_complex	0,0000001

Rappelons également que le modèle est le suivant:

$$P(T \le t) = 1 - \left(\frac{\alpha}{\alpha + B(t)} \right)^{r} \quad \text{avec} \quad B(t) = \sum_{i=1}^{t} \left[i^{c} - (i-1)^{c} \right] \exp(\beta x(i))$$

r et α sont les paramètres du modèle Gamma suivi par le paramètre λ

x(i) est le vecteur des covariés au temps t ;

c est le paramètre de forme de la loi de Weibull ;

β sont les effets des covariés.

Nous constatons que le paramètre c est inférieur à 1 ce qui signifie que le taux de hasard est décroissant (le taux instantané de non réponse en t conditionnel à la survie en t est décroissant), ce qui est cohérent avec l'analyse de la figure 5.24 : plus on avance dans le questionnaire, plus la probabilité d'arrêt est faible.

Les coefficients du type de question et du nombre de questions virales avant montrent que si la question est virale ou plus le nombre de questions virales avant est important, plus la probabilité cumulée d'arrêt à la question est forte. Notons également que les coefficients pour le numéro de la question ou la complexité de la question sont très faibles.

5.4.2.2 Modèle discret de non-réponse aux questions : :

Analyse critique du modèle Weibull Gamma :

Nous avons constaté les limites du modèles Weibull Gamma avec covariés dans la partie 5.4.2.1 (ce modèle sous estime la probabilité d'arrêt de certaines questions).
Le modèle Weibull Gamma avec 4 covariés que nous avons développé dans la partie précédente est de plus un modèle représentant un phénomène continu. Cependant, on peut analyser les non réponses aux questions comme un phénomène discret, dans lequel chaque individu répond de manière binaire (réponse/non réponse) à chaque question.
Nous avons donc testé un autre modèle de type discret pour modéliser le nombre de non-réponses par question. Ce modèle est de type béta-géométrique.

Modèle de type Béta Géométrique :

L'objectif ici est de modéliser le probabilité d'arrêt à la question N selon qu'elle soit virale ou de qualification.
Désignons par :

- p_1 : la probabilité d'un individu de répondre à une question virale
- p_2 : la probabilité d'un individu de répondre à une question de qualification
- n_1 : le nombre de questions virales auxquelles l'individu a déjà répondu lorsqu'on lui pose la question N
- n_2 : le nombre de questions de qualification auxquelles l'individu a déjà répondu par l'individu lorsqu'on lui pose la question N

Le fait de répondre ou ne pas répondre à une séquence de questions correspond à une séquence d'essais de Bernoulli (pour chaque question, chaque joueur a une probabilité p de répondre, qui sera égale à 0 quand il ne répond pas et 1 lorsqu'il répond) ; la distribution de Bernoulli est la fonction de densité d'une variable aléatoire discrète X prenant uniquement les valeurs 0 ou 1) jusqu'à la question où le joueur s'arrête. On peut donc supposer que la probabilité qu'un individu s'arrête à la question N suit une loi de type géométrique suivante :

$$\text{Proba (arrêt)} = p_1^{n1} \times p_2^{n2} \times (1-p_1) \text{ si la question N est virale} \qquad \text{(Equation 5.1)}$$

$$\text{Proba (arrêt)} = p_1^{n1} \times p_2^{n2} \times (1-p_2) \text{ si la question N est une question de qualification} \qquad \text{(Equation 5.2)}$$

Les lois Géométriques sont souvent utilisées pour modéliser en marketing des probabilités d'achat de produit (Hardie et Fader 2001 utilisent ce modèle pour représenter les achats et réachats de disques (données sur le site CD-now)), ainsi qu'en production pour étudier les problèmes de files d'attente (Trivedi, 2002).

Désignons par

- $q_1=1-p_1$ la probabilité d'arrêt du joueur à une question virale
- $q_2=1-p_2$ la probabilité d'arrêt du joueur à une question de qualification

En exprimant l'équation 5 .1 et 5.2 en fonction de q2 et q1 on obtient :

Equation 5.1(question virale) :

Proba (arrêt) = $(1-q_1)^{n1}\ (1-q_2)^{n2}\ q_1$

Equation 5.2(question de qualification) :

Proba (arrêt) = $(1-q_1)^{n1}\ (1-q_2)^{n2}\ q_2$

Pour calculer la probabilité que x joueurs s'arrêtent à la question N, nous introduisons dans notre modèle une hétérogénéité entre les individus pour la probabilité q_1 et pour la probabilité q_2 . L'hypothèse de notre modèle est que q_1 suit une loi Béta de paramètres alpha1 et béta1, et q_2 suit aussi une loi Béta de paramètres alpha2 et béta2. On modélise par conséquent la probabilité que X joueurs s'arrêtent à la question N par une loi de type Béta- Géométrique (Harder et Fardie, 2001). Dans les équations 5.1 et 5.2, q_1 et q_2 suivent par conséquent chacun une loi Béta. Ceci repose sur l'hypothèse selon laquelle la manière dont les joueurs répondent aux questions de qualification est indépendante de cette dont ils répondent aux questions virales. On peut en effet avoir des joueurs prêts à répondre à des questions de qualification mais peu enclins à fournir des adresses e-mail, ou encore des joueurs prêts à fournir des adresses virales mais n'ayant pas envie de fournir des informations qualifiées les concernant.

Rappel de la distribution Béta :

La distribution Béta est souvent utilisée en marketing pour modéliser des lois mixtes représentant des probabilité d'achat (Lilien, Kotler et Moorthy, p. 602). Elle se caractérise par :

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}X^{\alpha-1}(1-X)^{\beta-1}$$

$$0\leq X\leq 1$$

L'espérance et la variance de la loi sont :

$$E(X)=\frac{\alpha}{\alpha+\beta}$$

$$V(X)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Les paramètres du modèle Béta-Géométrique sont estimés en maximisant le logarithme de la vraisemblance, qui s'écrit :

LL = $n_1 \log (1-q_1) + n_2 \log (1-q_2) + \log q_1$ si la question N est virale

$+ \log (q_2)$ si la question N est une question de qualification

avec q_1 suit une loi Béta (α_1,β_1),

q_2 suit une loi Béta (α_2,β_2),

Calcul des paramètres du modèle de type Béta-Géométrique :

Les tableaux ci-dessous fournissent les formules utilisées sous excel ainsi que les résultats obtenus pour les 41 questions.

Plusieurs éléments du modèle apparaissent dans les formules et les résultats :

1. Le nombre de questions virales répondues précédemment (n_1) et le nombre de questions de qualification répondues précédemment (n_2) sont des chiffres décimaux, car ils sont calculés à partir d'un fichier par individu reconstituant pour chaque individu et chaque question le nombre de questions précédentes de qualification et virales. C'est donc la moyenne du nombre de questions virales ou de qualification sur les individus ayant répondu à la question. Le chiffre est décimal en raison des questions mixtes, pour lesquelles certains individus sont soumis à une question virale et d'autres à une question de qualification (par exemple le nombre de questions virales précédentes vaut 9,56 pour la question Q14 car cette question est envoyée à un mélange de joueurs qui pour certains d'entre eux ont répondu à la question Q13 virale, et pour d'autres à la question Q13 de qualification).

2. Les questions mixtes ont été traitées en séparant les questions virales et de qualification (pour la question 13 par exemple, nous avons dans le fichier de données une ligne pour la question Q13 virale et une ligne pour la question Q13 de qualification).

3. Les paramètres α_1, β_1, α_2, β_2 sont estimés dans le modèle en maximisant la vraisemblance.

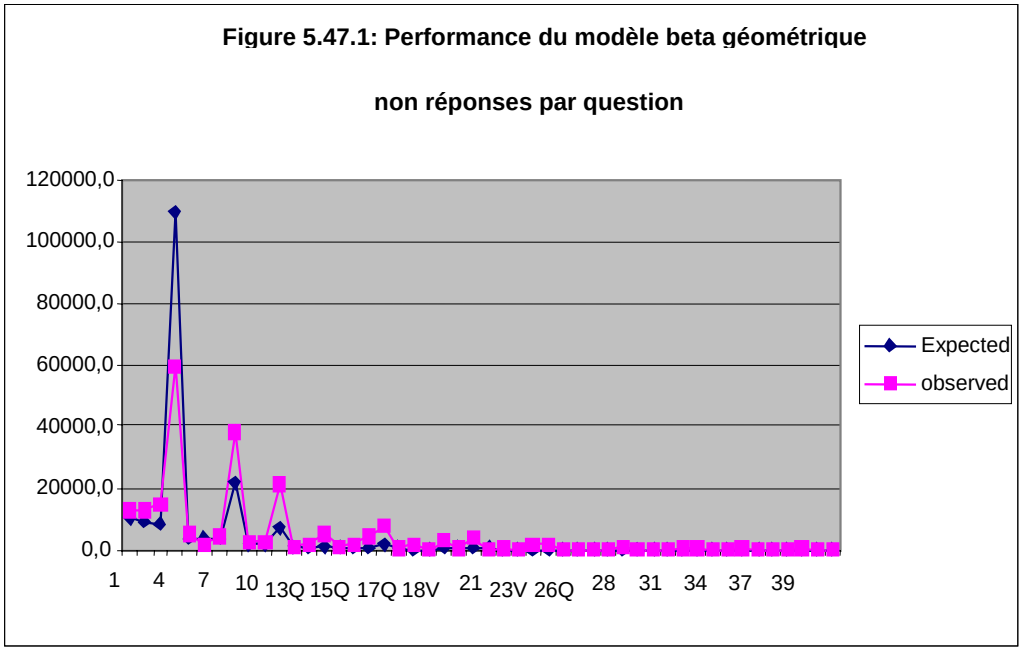
Tableau 5.55 : Modèle de type Béta-Géométrique: Formules

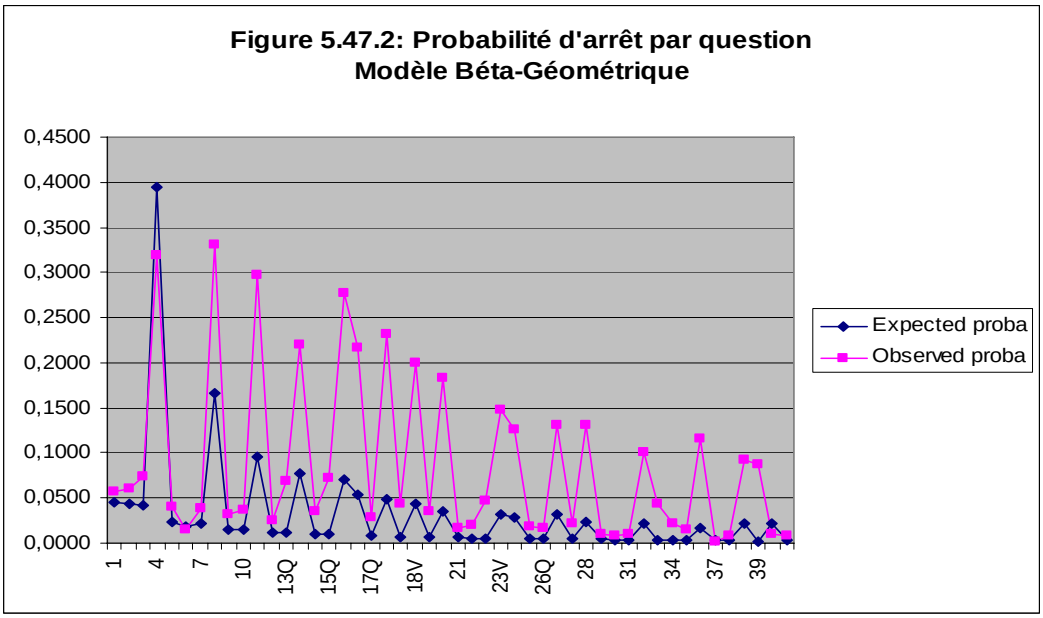
alpha2	0,001027644						
beta2	0,000048452						
alpha1	0,658282328				LL =	=SOMME(J6:J53	
beta1	2,440944827						
Q	ion réponses	ion	P(T=x)	LL	n2	n1	
1	13111	1	=((1-LOI.BETA(F6;F$3;F$4;0;227879))^L6)*((1-LOI.BETA(F6;F$1;F$2;0;227879))^K6)*(LOI.BETA(F6;F$1;F$2;0;227879))	=F6*LN(I6)	0	0	
2	13146	1	=((1-LOI.BETA(F7;F$3;F$4;0;227879))^L7)*((1-LOI.BETA(F7;F$1;F$2;0;227879))^K7)*(LOI.BETA(F7;F$1;F$2;0;227879))	=F7*LN(I7)	1	0	
3	14729	1	=((1-LOI.BETA(F8;F$3;F$4;0;227879))^L8)*((1-LOI.BETA(F8;F$1;F$2;0;227879))^K8)*(LOI.BETA(F8;F$1;F$2;0;227879))	=F8*LN(I8)	2	0	
4	59518	2	=((1-LOI.BETA(F9;F$3;F$4;0;227879))^L9)*((1-LOI.BETA(F9;F$1;F$2;0;227879))^K9)*(LOI.BETA(F9;F$3;F$4;0;227879))	=F9*LN(I13)	3	0	
5	5216	1	=((1-LOI.BETA(F10;F$3;F$4;0;227879))^L10)*((1-LOI.BETA(F10;F$1;F$2;0;227879))^K10)*(LOI.BETA(F10;F$1;F$2;0;227879))	=F10*LN(I10)	3	1	
6	1896	1	=((1-LOI.BETA(F11;F$3;F$4;0;227879))^L11)*((1-LOI.BETA(F11;F$1;F$2;0;227879))^K11)*(LOI.BETA(F11;F$1;F$2;0;227879))	=F11*LN(I11)	4	1	
7	4546	1	=((1-LOI.BETA(F12;F$3;F$4;0;227879))^L12)*((1-LOI.BETA(F12;F$1;F$2;0;227879))^K12)*(LOI.BETA(F12;F$1;F$2;0;227879))	=F12*LN(I12)	5	1	
8	38349	2	=((1-LOI.BETA(F13;F$3;F$4;0;227879))^L13)*((1-LOI.BETA(F13;F$1;F$2;0;227879))^K13)*(LOI.BETA(F13;F$3;F$4;0;227879))	=F13*LN(I13)	6	1	
9	2431	1	=((1-LOI.BETA(F14;F$3;F$4;0;227879))^L14)*((1-LOI.BETA(F14;F$1;F$2;0;227879))^K14)*(LOI.BETA(F14;F$1;F$2;0;227879))	=F14*LN(I14)	6	2	
10	2810	1	=((1-LOI.BETA(F15;F$3;F$4;0;227879))^L15)*((1-LOI.BETA(F15;F$1;F$2;0;227879))^K15)*(LOI.BETA(F15;F$1;F$2;0;227879))	=F15*LN(I15)	7	2	
11	21460	2	=((1-LOI.BETA(F16;F$3;F$4;0;227879))^L16)*((1-LOI.BETA(F16;F$1;F$2;0;227879))^K16)*(LOI.BETA(F16;F$3;F$4;0;227879))	=F16*LN(I16)	8	2	
12	1305	1	=((1-LOI.BETA(F17;F$3;F$4;0;227879))^L17)*((1-LOI.BETA(F17;F$1;F$2;0;227879))^K17)*(LOI.BETA(F17;F$1;F$2;0;227879))	=F17*LN(I17)	8	3	
13Q	1679	1	=((1-LOI.BETA(F18;F$3;F$4;0;227879))^L18)*((1-LOI.BETA(F18;F$1;F$2;0;227879))^K18)*(LOI.BETA(F18;F$1;F$2;0;227879))	=F18*LN(I18)	9,1158547153	2,884145284	
13V	5457	2	=((1-LOI.BETA(F19;F$3;F$4;0;227879))^L19)*((1-LOI.BETA(F19;F$1;F$2;0;227879))^K19)*(LOI.BETA(F19;F$3;F$4;0;227879))	=F19*LN(I19)	9	3	
14	1498	1	=((1-LOI.BETA(F20;F$3;F$4;0;227879))^L20)*((1-LOI.BETA(F20;F$1;F$2;0;227879))^K20)*(LOI.BETA(F20;F$1;F$2;0;227879))	=F20*LN(I20)	9,5552057345	3,44	
15Q	=E21	1	=((1-LOI.BETA(F21;F$3;F$4;0;227879))^L21)*((1-LOI.BETA(F21;F$1;F$2;0;227879))^K21)*(LOI.BETA(F21;F$1;F$2;0;227879))	=F21*LN(I21)	10,375021125	3,62	
15V	=E22	2	=((1-LOI.BETA(F22;F$3;F$4;0;227879))^L22)*((1-LOI.BETA(F22;F$1;F$2;0;227879))^K22)*(LOI.BETA(F22;F$3;F$4;0;227879))	=F22*LN(I22)	10,840258205	3,16	
16	=E23	2	=((1-LOI.BETA(F23;F$3;F$4;0;227879))^L23)*((1-LOI.BETA(F23;F$1;F$2;0;227879))^K23)*(LOI.BETA(F23;F$3;F$4;0;227879))	=F23*LN(I23)	11,180609133	3,82	
17Q	=E24	1	=((1-LOI.BETA(F24;F$3;F$4;0;227879))^L24)*((1-LOI.BETA(F24;F$1;F$2;0;227879))^K24)*(LOI.BETA(F24;F$1;F$2;0;227879))	=F24*LN(I24)	11,311959087	4,69	
17V	=E25	2	=((1-LOI.BETA(F25;F$3;F$4;0;227879))^L25)*((1-LOI.BETA(F25;F$1;F$2;0;227879))^K25)*(LOI.BETA(F25;F$3;F$4;0;227879))	=F25*LN(I25)	10,831472230	5,17	
18Q	=E26	1	=((1-LOI.BETA(F26;F$3;F$4;0;227879))^L26)*((1-LOI.BETA(F26;F$1;F$2;0;227879))^K26)*(LOI.BETA(F26;F$1;F$2;0;227879))	=F26*LN(I26)	11,885615749	5,11	
18V	=E27	2	=((1-LOI.BETA(F27;F$3;F$4;0;227879))^L27)*((1-LOI.BETA(F27;F$1;F$2;0;227879))^K27)*(LOI.BETA(F27;F$3;F$4;0;227879))	=F27*LN(I27)	11,922610890	5,08	
19	=E28	1	=((1-LOI.BETA(F28;F$3;F$4;0;227879))^L28)*((1-LOI.BETA(F28;F$1;F$2;0;227879))^K28)*(LOI.BETA(F28;F$1;F$2;0;227879))	=F28*LN(I28)	12,234937965	5,77	
20	=E29	2	=((1-LOI.BETA(F29;F$3;F$4;0;227879))^L29)*((1-LOI.BETA(F29;F$1;F$2;0;227879))^K29)*(LOI.BETA(F29;F$3;F$4;0;227879))	=F29*LN(I29)	13,237860232	5,76	
21	=E30	1	=((1-LOI.BETA(F30;F$3;F$4;0;227879))^L30)*((1-LOI.BETA(F30;F$1;F$2;0;227879))^K30)*(LOI.BETA(F30;F$1;F$2;0;227879))	=F30*LN(I30)	13,214560337	6,79	
22	=E31	1	=((1-LOI.BETA(F31;F$3;F$4;0;227879))^L31)*((1-LOI.BETA(F31;F$1;F$2;0;227879))^K31)*(LOI.BETA(F31;F$1;F$2;0;227879))	=F31*LN(I31)	14,215294043	6,78	
23Q	=E32	1	=((1-LOI.BETA(F32;F$3;F$4;0;227879))^L32)*((1-LOI.BETA(F32;F$1;F$2;0;227879))^K32)*(LOI.BETA(F32;F$1;F$2;0;227879))	=F32*LN(I32)	15,449068164	6,55	
23V	=E33	2	=((1-LOI.BETA(F33;F$3;F$4;0;227879))^L33)*((1-LOI.BETA(F33;F$1;F$2;0;227879))^K33)*(LOI.BETA(F33;F$3;F$4;0;227879))	=F33*LN(I33)	15,103187411	6,9	
24	=E34	2	=((1-LOI.BETA(F34;F$3;F$4;0;227879))^L34)*((1-LOI.BETA(F34;F$1;F$2;0;227879))^K34)*(LOI.BETA(F34;F$3;F$4;0;227879))	=F34*LN(I34)	15,509220210	7,49	
25	=E35	1	=((1-LOI.BETA(F35;F$3;F$4;0;227879))^L35)*((1-LOI.BETA(F35;F$1;F$2;0;227879))^K35)*(LOI.BETA(F35;F$1;F$2;0;227879))	=F35*LN(I35)	15,487246401	8,51	
26Q	=E36	1	=((1-LOI.BETA(F36;F$3;F$4;0;227879))^L36)*((1-LOI.BETA(F36;F$1;F$2;0;227879))^K36)*(LOI.BETA(F36;F$1;F$2;0;227879))	=F36*LN(I36)	15,487246401	8,51	
26V	=E37	2	=((1-LOI.BETA(F37;F$3;F$4;0;227879))^L37)*((1-LOI.BETA(F37;F$1;F$2;0;227879))^K37)*(LOI.BETA(F37;F$3;F$4;0;227879))	=F37*LN(I37)	16,511001708	8,49	
27	=E38	1	=((1-LOI.BETA(F38;F$3;F$4;0;227879))^L38)*((1-LOI.BETA(F38;F$1;F$2;0;227879))^K38)*(LOI.BETA(F38;F$1;F$2;0;227879))	=F38*LN(I38)	17,356115746	8,64	
28	=E39	2	=((1-LOI.BETA(F39;F$3;F$4;0;227879))^L39)*((1-LOI.BETA(F39;F$1;F$2;0;227879))^K39)*(LOI.BETA(F39;F$3;F$4;0;227879))	=F39*LN(I39)	18,369090345	8,63	
29	=E40	1	=((1-LOI.BETA(F40;F$3;F$4;0;227879))^L40)*((1-LOI.BETA(F40;F$1;F$2;0;227879))^K40)*(LOI.BETA(F40;F$1;F$2;0;227879))	=F40*LN(I40)	18,339722337	9,66	
30	=E41	1	=((1-LOI.BETA(F41;F$3;F$4;0;227879))^L41)*((1-LOI.BETA(F41;F$1;F$2;0;227879))^K41)*(LOI.BETA(F41;F$1;F$2;0;227879))	=F41*LN(I41)	19,341830338	9,66	
31	=E42	1	=((1-LOI.BETA(F42;F$3;F$4;0;227879))^L42)*((1-LOI.BETA(F42;F$1;F$2;0;227879))^K42)*(LOI.BETA(F42;F$1;F$2;0;227879))	=F42*LN(I42)	20,343891886	9,66	
32	=E43	2	=((1-LOI.BETA(F43;F$3;F$4;0;227879))^L43)*((1-LOI.BETA(F43;F$1;F$2;0;227879))^K43)*(LOI.BETA(F43;F$3;F$4;0;227879))	=F43*LN(I43)	21,369600935	9,63	
33	=E44	1	=((1-LOI.BETA(F44;F$3;F$4;0;227879))^L44)*((1-LOI.BETA(F44;F$1;F$2;0;227879))^K44)*(LOI.BETA(F44;F$1;F$2;0;227879))	=F44*LN(I44)	21,494798439	10,51	
34	=E45	1	=((1-LOI.BETA(F45;F$3;F$4;0;227879))^L45)*((1-LOI.BETA(F45;F$1;F$2;0;227879))^K45)*(LOI.BETA(F45;F$1;F$2;0;227879))	=F45*LN(I45)	22,081379310	10,92	
35	=E46	1	=((1-LOI.BETA(F46;F$3;F$4;0;227879))^L46)*((1-LOI.BETA(F46;F$1;F$2;0;227879))^K46)*(LOI.BETA(F46;F$1;F$2;0;227879))	=F46*LN(I46)	23,08	10,92	
36	=E47	2	=((1-LOI.BETA(F47;F$3;F$4;0;227879))^L47)*((1-LOI.BETA(F47;F$1;F$2;0;227879))^K47)*(LOI.BETA(F47;F$3;F$4;0;227879))	=F47*LN(I47)	24,111702895	10,89	
37	=E48	1	=((1-LOI.BETA(F48;F$3;F$4;0;227879))^L48)*((1-LOI.BETA(F48;F$1;F$2;0;227879))^K48)*(LOI.BETA(F48;F$1;F$2;0;227879))	=F48*LN(I48)	24,085087972	11,91	
38Q	=E49	1	=((1-LOI.BETA(F49;F$3;F$4;0;227879))^L49)*((1-LOI.BETA(F49;F$1;F$2;0;227879))^K49)*(LOI.BETA(F49;F$1;F$2;0;227879))	=F49*LN(I49)	25,159274903	11,84	
38V	=E50	2	=((1-LOI.BETA(F50;F$3;F$4;0;227879))^L50)*((1-LOI.BETA(F50;F$1;F$2;0;227879))^K50)*(LOI.BETA(F50;F$3;F$4;0;227879))	=F50*LN(I50)	24,703669724	12,3	
39	=E51	1	=((1-LOI.BETA(F51;F$3;F$4;0;227879))^L51)*((1-LOI.BETA(F51;F$1;F$2;0;227879))^K51)*(LOI.BETA(F51;F$1;F$2;0;227879))	=F51*LN(I51)	25,925286500	12,07	
40	=E52	2	=((1-LOI.BETA(F52;F$3;F$4;0;227879))^L52)*((1-LOI.BETA(F52;F$1;F$2;0;227879))^K52)*(LOI.BETA(F52;F$3;F$4;0;227879))	=F52*LN(I52)	26,945760496	12,05	
41	=E53	1	=((1-LOI.BETA(F53;F$3;F$4;0;227879))^L53)*((1-LOI.BETA(F53;F$1;F$2;0;227879))^K53)*(LOI.BETA(F53;F$1;F$2;0;227879))	=F53*LN(I53)	26,945015645	13,05	

Tableau 5.56 : Modèle de type Béta-Géométrique (Résultats)

				alpha2	0,001027644							
	227879			beta2	0,00004845							
				alpha1	0,66			LL =	-554500,21			
				beta1	2,44	Observed						
Nbre envois	Nbre réponse		Q	Envois	Non réponses	le Non réponse	Type question	P(T=x)	LL	n2	n1	E(n_{1x})
227879	214768	1	1	13111	13111	0,0575	1	0,0449	-40688,3655	0,00	0	10231,2
214768	201622	2	2	13146	13146	0,0612	1	0,0429	-41400,8281	1,00	0	9209,6
201622	186893	3	3	14729	14729	0,0731	1	0,0410	-47061,1255	2,00	0	8258,6
186893	127375	4	4	59518	59518	0,3185	2	0,5882	-99227,6863	3,00	0	109930,2
127375	122159	5	5	5216	5216	0,0409	1	0,0330	-17799,7189	3,00	1	4198,0
122159	120263	6	6	1896	1896	0,0155	1	0,0343	-6396,0796	4,00	1	4186,5
120263	115717	7	7	4546	4546	0,0378	1	0,0305	-15859,8135	5,00	1	3672,7
115717	77368	8	8	38349	38349	0,3314	2	0,1888	-63934,9868	6,00	1	21844,7
77368	74937	9	9	2431	2431	0,0314	1	0,0278	-8705,4606	6,00	2	2154,5
74937	72127	10	10	2810	2810	0,0375	1	0,0260	-10250,2948	7,00	2	1952,0
72127	50667	11	11	21460	21460	0,2975	2	0,1010	-49192,6934	8,00	2	7287,3
50667	49362	12	12	1305	1305	0,0258	1	0,0255	-4788,6785	8,00	3	1291,5
24551	22872	13	13Q	1679	1679	0,0684	1	0,0236	-6293,2165	9,12	2,88	578,4
24811	19354	13	13V	5457	5457	0,2199	2	0,0630	-15088,6732	9,00	3	1562,5
42226	40728	14	14	1498	1498	0,0355	1	0,0226	-5680,0090	9,56	3,44	952,5
25501	23668	15	15Q	1833	1833	0,0719	1	0,0206	-7113,1338	10,38	3,62	526,3
15227	10999	15	15V	4228	4228	0,2777	2	0,0523	-12477,9116	10,84	3,16	796,0
34667	27179	16	16	7488	7488	0,2160	2	0,0510	-22282,1748	11,18	3,82	1768,4
18325	17794	17	17Q	531	531	0,0290	1	0,0225	-2014,0231	11,31	4,69	412,9
8854	6806	17	17V	2048	2048	0,2313	2	0,0328	-7001,1858	10,83	5,17	290,1
7273	6959	18	18Q	314	314	0,0432	1	0,0228	-1186,9930	11,89	5,11	166,0
17327	13865	18	18V	3462	3462	0,1998	2	0,0363	-11480,0134	11,92	5,08	628,9
20824	20105	19	19	719	719	0,0345	1	0,0198	-2818,7567	12,23	5,77	413,0
20105	16413	20	20	3692	3692	0,1836	2	0,0316	-12757,2243	13,24	5,76	634,8
16413	16126	21	21	287	287	0,0175	1	0,0208	-1111,7125	13,21	6,79	341,1
16126	15797	22	22	329	329	0,0204	1	0,0196	-1294,4257	14,22	6,78	315,4
4113	3917	23	23Q	196	196	0,0477	1	0,0195	-771,3136	15,45	6,55	80,4
11624	9914	23	23V	1710	1710	0,1471	2	0,0220	-6524,1351	15,10	6,9	256,1
13831	12093	24	24	1738	1738	0,1257	2	0,0207	-6737,2269	15,51	7,49	286,6
12093	11879	25	25	214	214	0,0177	1	0,0186	-852,2877	15,49	8,51	225,4
10130	9953	26	26Q	177	177	0,0175	1	0,0190	-701,4460	15,49	8,51	192,5
1749	1520	26	26V	229	229	0,1309	2	0,0080	-1105,4525	16,51	8,49	14,0
11473	11226	27	27	247	247	0,0215	1	0,0168	-1009,6285	17,36	8,64	192,5
11226	9751	28	28	1475	1475	0,1314	2	0,0160	-6096,2079	18,37	8,63	180,0
9751	9652	29	29	99	99	0,0102	1	0,0172	-401,9442	18,34	9,66	168,2
9652	9572	30	30	80	80	0,0083	1	0,0167	-327,2746	19,34	9,66	161,4
9572	9471	31	31	101	101	0,0106	1	0,0157	-419,4672	20,34	9,66	150,4
9471	8520	32	32	951	951	0,1004	2	0,0116	-4234,4129	21,37	9,63	110,3
8520	8152	33	33	368	368	0,0432	1	0,0125	-1614,0455	21,49	10,51	106,1
8152	7975	34	34	177	177	0,0217	1	0,0135	-762,1813	22,08	10,92	109,9
7975	7855	35	35	120	120	0,0150	1	0,0134	-517,0695	23,08	10,92	107,2
7855	6947	36	36	908	908	0,1156	2	0,0095	-4229,1234	24,11	10,89	74,5
6947	6934	37	37	13	13	0,0019	1	0,0143	-55,1910	24,09	11,91	99,5
5733	5682	38	38Q	51	51	0,0089	1	0,0129	-221,7371	25,16	11,84	74,2
1201	1090	38	38V	111	111	0,0924	2	0,0035	-628,2081	24,70	12,3	4,2
6772	6179	39	39	593	593	0,0876	1	0,0086	-2822,6574	25,93	12,07	58,0
6179	6121	40	40	58	58	0,0094	2	0,0022	-355,6466	26,95	12,05	13,4
6121	6074	41	41	47	47	0,0077	1	0,0119	-208,3645	26,95	13,05	72,7

La performance du modèle est fournie par les figure 5.47.1 et 5.47.2 ci-dessous :





On constate que le modèle suit globalement les données observées, notamment la hausse du nombre de non réponses lors des questions virales et mixtes. La figure 5.47.2 des probabilités d'arrêt montre cependant des écarts entre les probabilités d'arrêt observées et celles estimées par le modèle, notamment à partir de la 8ème question.

Si on compare le modèle Béta Géométrique avec le modèle Weibull-Gamma avec 4 covariés que nous avons développé dans la partie précédente, nous constatons (cf Tableau 5.57) que le modèle WG + 4 covariés donne une meilleure vraisemblance et qu'elle est significativement meilleure (67833>7,81).

Tableau 5.57: Comparaison des modèles WG + 4 covariés et Béta Géométrique

		nbr paramètres
LL Béta Géométrique	-554500	4
LL WG + 4 Covariés	-520584	7
LR	67833,37	
Chi-2 (0,95,3)	7,81	

Le modèle Béta Géométrique présente cependant plusieurs avantages intéressants: il tient tout d'abord bien compte du comportement des individus déjà constaté précédemment (figure 5.38), à savoir que plus on avance dans les questions virales, plus la probabilité de réponse est

forte (donc plus la probabilité de non réponse est faible), comme on peut le constater dans le tableau 5.56 des résultats.
En revanche, le modèle présente des écarts avec les données observées. Il surestime en particulier le taux de non réponse à la première question virale et sous-estime ensuite les non réponses aux questions virales suivantes. La courbe 5.47.2 montre que la faiblesse du modèle béta géométrique est l'estimation de la probabilité d'arrêt aux questions virales.
Afin d'améliorer le modèle, nous avons donc modifié dans le modèle le calcul de la probabilité d'arrêt aux questions virales. Etant donné la sous-estimation des probabilités d'arrêt aux questions virales, nous avons introduit un coefficient correcteur, uniquement sur les questions virales, ce coefficient multiplicateur accentuant l'impact du nombre de questions virales auquel le joueur a déjà été soumis. Le coefficient multiplicatif que nous avons choisi est le numéro de la question virale posée. Nous émettons en effet l'hypothèse selon laquelle ce coefficient exprime le fait que le joueur a un stock d'adresses e-mail limité et que plus il répond à des questions virales, plus son stock diminue et sa probabilité d'arrêt à la question virale posée augmente. Nous captons ainsi deux effets inverses dans notre modèle : un effet « stock d'adresses virales » augmentant la probabilité d'arrêt au fur et à mesure des questions virales, un effet « hétérogénéité » des joueurs, les joueurs qui continuent à jouer en répondant à un nombre élevé de questions virales ayant eux une probabilité d'arrêt plus faible que les joueurs qui s'arrêtent aux premières questions virales.
La probabilité d'arrêt à une question virale est donc la suivante :

Equation 5.3 (question virale) :

Proba (arrêt) = $(1-q_1)\,(1-2q_1)\,(1-3q_1)\,\ldots(1-n_1q_1)\,(1-q_2)^{n2}\,(n_1+1)\,q_1$

avec

- n_1 : nombre de questions virales auxquelles l'individu a déjà répondu lorsqu'on lui pose la question N
- (n_1+1) correspond au numéro de la question virale à laquelle est soumise le joueur (ce numéro est déterminé uniquement sur l'ensemble des questions virales).
- n_2 : nombre de questions de qualification auxquelles l'individu a déjà répondu par l'individu lorsqu'on lui pose la question N
- q_1 la probabilité d'arrêt du joueur à une question virale (q_1 suit une loi Béta)
- q_2 la probabilité d'arrêt du joueur à une question de qualification (q_2 suit une loi Béta)

- Une contrainte du modèle est que $n_1+1 \leq \frac{1}{(1-q_1)(1-2q_1)...(1-n_1q_1)(1-q_2)^{n_2}q_1}$

D'autre part, nous avons traité spécifiquement la probabilité de réponse à la première question virale, dont la probabilité d'arrêt était fortement sur-estimée dans le modèle précédent.. Nous avons pour cela émis l'hypothèse selon laquelle une partie des individus soumis à la première question virale sont réfractaires lorsqu'on leur demande de fournir une information (adresse e-mail) sur une autre personne. Cette hypothèse se justifie par le comportement actuel de certains internautes réfractaires à fournir des informations qu'ils considèrent comme confidentielles ou personnelles.

Pour tenir compte de cette hypothèse, nous avons donc ajouté, uniquement pour la première question virale, un coefficient spécifique dans l'équation de la probabilité d'arrêt.

La probabilité d'arrêt à une question de qualification n'est quant à elle pas modifiée par rapport au premier modèle Béta Géométrique.

Ce nouveau modèle sera nommé ultérieurement modèle Béta Géométrique 2, le premier modèle Béta Géométrique sera nommé Béta Géométrique 1.

Les formules du tableau excel sont modifiées de la manière suivante (les formules devenant de plus en plus longues au fur et à mesure qu'on avance dans le questionnaire, ce tableau montre uniquement les 20 premières questions).

Tableau 5.58 : formules modèle Béta Géométrique 2

P(T=x)	LL	ı questi
=((1-LOI.BETA(F10;F$5;F$6;0;227879))^O10)*((1-LOI.BETA(F10;F$3;F$4;0;227879))^N10)*(LOI.BETA(F10;F$3;F$4;0;227879))	=F10*LN(I10)	1
=((1-LOI.BETA(F11;F$5;F$6;0;227879))^O11)*((1-LOI.BETA(F11;F$3;F$4;0;227879))^N11)*(LOI.BETA(F11;F$3;F$4;0;227879))	=F11*LN(I11)	2
=((1-LOI.BETA(F12;F$5;F$6;0;227879))^O12)*((1-LOI.BETA(F12;F$3;F$4;0;227879))^N12)*(LOI.BETA(F12;F$3;F$4;0;227879))	=F12*LN(I12)	3
=F$2*K13*((1-LOI.BETA(F13;F$5;F$6;0;227879))^O13)*((1-LOI.BETA(F13;F$3;F$4;0;227879))^N13)*(LOI.BETA(F13;F$5;F$6;0;227879))	=F13*LN(I17)	4
=((1-LOI.BETA(F14;F$5;F$6;0;227879))^O14)*((1-LOI.BETA(F14;F$3;F$4;0;227879))^N14)*(LOI.BETA(F14;F$3;F$4;0;227879))	=F14*LN(I14)	5
=((1-LOI.BETA(F15;F$5;F$6;0;227879))^O15)*((1-LOI.BETA(F15;F$3;F$4;0;227879))^N15)*(LOI.BETA(F15;F$3;F$4;0;227879))	=F15*LN(I15)	6
=((1-LOI.BETA(F16;F$5;F$6;0;227879))^O16)*((1-LOI.BETA(F16;F$3;F$4;0;227879))^N16)*(LOI.BETA(F16;F$3;F$4;0;227879))	=F16*LN(I16)	7
=K17*(1-4*LOI.BETA(F17;F$5;F$6;0;227879))*((1-LOI.BETA(F17;F$3;F$4;0;227879))^N17)*(LOI.BETA(F17;F$5;F$6;0;227879))	=F17*LN(I17)	8
=((1-LOI.BETA(F18;F$5;F$6;0;227879))^O18)*((1-LOI.BETA(F18;F$3;F$4;0;227879))^N18)*(LOI.BETA(F18;F$3;F$4;0;227879))	=F18*LN(I18)	9
=((1-LOI.BETA(F19;F$5;F$6;0;227879))^O19)*((1-LOI.BETA(F19;F$3;F$4;0;227879))^N19)*(LOI.BETA(F19;F$3;F$4;0;227879))	=F19*LN(I19)	10
=K20*(1-4*LOI.BETA(F20;F$5;F$6;0;227879))*(1-8*LOI.BETA(F20;F$5;F$6;0;227879))*((1-LOI.BETA(F20;F$3;F$4;0;227879))^N20)*(LOI.BETA(F20;F$5;F$6;0;227879))	=F20*LN(I20)	11
=((1-LOI.BETA(F21;F$5;F$6;0;227879))^O21)*((1-LOI.BETA(F21;F$3;F$4;0;227879))^N21)*(LOI.BETA(F21;F$3;F$4;0;227879))	=F21*LN(I21)	12
=((1-LOI.BETA(F22;F$5;F$6;0;227879))^O22)*((1-LOI.BETA(F22;F$3;F$4;0;227879))^N22)*(LOI.BETA(F22;F$3;F$4;0;227879))	=F22*LN(I22)	13
=K23*(1-4*LOI.BETA(F23;F$5;F$6;0;227879))*(1-8*LOI.BETA(F23;F$5;F$6;0;227879))*(1-11*LOI.BETA(F23;F$5;F$6;0;227879))*((1-LOI.BETA(F23;F$3;F$4;0;227879))^N23)*(L(	=F23*LN(I23)	13
=((1-LOI.BETA(F24;F$5;F$6;0;227879))^O24)*((1-LOI.BETA(F24;F$3;F$4;0;227879))^N24)*(LOI.BETA(F24;F$3;F$4;0;227879))	=F24*LN(I24)	14
=((1-LOI.BETA(F25;F$5;F$6;0;227879))^O25)*((1-LOI.BETA(F25;F$3;F$4;0;227879))^N25)*(LOI.BETA(F25;F$3;F$4;0;227879))	=F25*LN(I25)	15
=K26*(1-4*LOI.BETA(F26;F$5;F$6;0;227879))*(1-8*LOI.BETA(F26;F$5;F$6;0;227879))*(1-11*LOI.BETA(F26;F$5;F$6;0;227879))*(1-13*LOI.BETA(F26;F$5;F$6;0;227879))*((1-LC	=F26*LN(I26)	15
=K27*(1-4*LOI.BETA(F27;F$5;F$6;0;227879))*(1-8*LOI.BETA(F27;F$5;F$6;0;227879))*(1-11*LOI.BETA(F27;F$5;F$6;0;227879))*(1-13*LOI.BETA(F27;F$5;F$6;0;227879))*(1-15*	=F27*LN(I27)	16
=((1-LOI.BETA(F28;F$5;F$6;0;227879))^O28)*((1-LOI.BETA(F28;F$3;F$4;0;227879))^N28)*(LOI.BETA(F28;F$3;F$4;0;227879))	=F28*LN(I28)	17
=K29*(1-4*LOI.BETA(F29;F$5;F$6;0;227879))*(1-8*LOI.BETA(F29;F$5;F$6;0;227879))*(1-11*LOI.BETA(F29;F$5;F$6;0;227879))*(1-13*LOI.BETA(F29;F$5;F$6;0;227879))*(1-15*	=F29*LN(I29)	17
=((1-LOI.BETA(F30;F$5;F$6;0;227879))^O30)*((1-LOI.BETA(F30;F$3;F$4;0;227879))^N30)*(LOI.BETA(F30;F$3;F$4;0;227879))	=F30*LN(I30)	18
=K31*(1-4*LOI.BETA(F31;F$5;F$6;0;227879))*(1-8*LOI.BETA(F31;F$5;F$6;0;227879))*(1-11*LOI.BETA(F31;F$5;F$6;0;227879))*(1-13*LOI.BETA(F31;F$5;F$6;0;227879))*(1-15*	=F31*LN(I31)	18
=((1-LOI.BETA(F32;F$5;F$6;0;227879))^O32)*((1-LOI.BETA(F32;F$3;F$4;0;227879))^N32)*(LOI.BETA(F32;F$3;F$4;0;227879))	=F32*LN(I32)	19
=K33*(1-4*LOI.BETA(F33;F$5;F$6;0;227879))*(1-8*LOI.BETA(F33;F$5;F$6;0;227879))*(1-11*LOI.BETA(F33;F$5;F$6;0;227879))*(1-13*LOI.BETA(F33;F$5;F$6;0;227879))*(1-15*	=F33*LN(I33)	20

Les résultats du modèle sont fournis ci-dessous, la performance étant présentée dans les figures 5.48.1 et 5.48.2.

Tableau 5.59 : Résultats modèle Béta Géométrique 2

alpha3	0,54
alpha2	0,00124
beta2	0,00006
alpha1	0,87
beta1	0,49

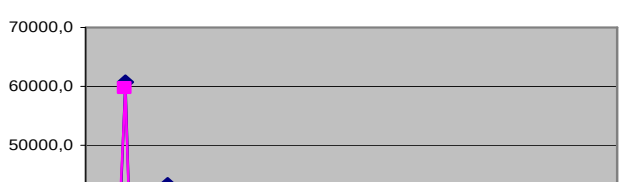
LL = -499904,89

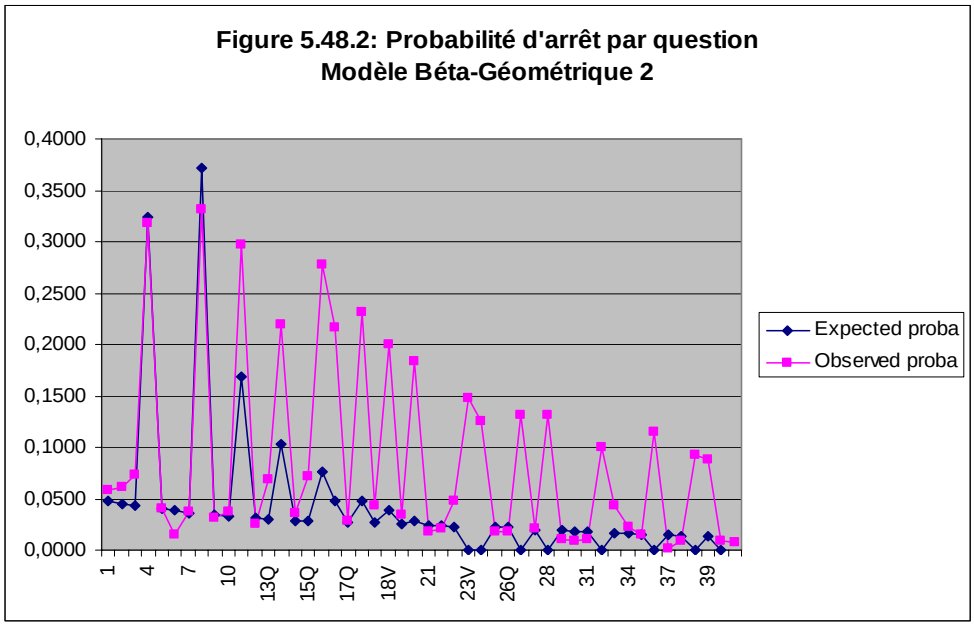
0

numéro numéro

Q	Non réponses	proba réelle Non réponse	P(T=x)	LL	no question	qualif	question virale	n2	n1	E(n_{1x})
1	13111	0,0575	0,0471	-40068,3699	1	1	0	0,00	0	10726,6
2	13146	0,0612	0,0449	-40809,1285	2	2	0	1,00	0	9633,6
3	14729	0,0731	0,0428	-46431,4072	3	3	0	2,00	0	8619,4
4	59518	0,3185	0,3243	-58843,1831	4	3	1	3,00	0	60611,0
5	5216	0,0409	0,0399	-16804,3812	5	4	1	3,00	1	5080,6
6	1896	0,0155	0,0384	-6179,3491	6	5	1	4,00	1	4693,5
7	4546	0,0378	0,0363	-15073,9452	7	6	1	5,00	1	4365,8
8	38349	0,3314	0,3721	-37914,1979	8	6	2	6,00	1	43055,3
9	2431	0,0314	0,0345	-8185,9719	9	7	2	6,00	2	2667,8
10	2810	0,0375	0,0328	-9604,7922	10	8	2	7,00	2	2456,1
11	21460	0,2975	0,1681	-38268,7829	11	8	3	8,00	2	12123,8
12	1305	0,0258	0,0314	-4517,1293	12	9	3	8,00	3	1590,2
13Q	1679	0,0684	0,0296	-5907,3677	13	10	3	9,12	2,88	727,9
13V	5457	0,2199	0,1028	-12414,9622	13	10	4	9,00	3	2550,4
14	1498	0,0355	0,0290	-5304,8538	14	11	4	9,56	3,44	1223,5
15Q	1833	0,0719	0,0277	-6573,9651	15	12	4	10,38	3,62	706,3
15V	4228	0,2777	0,0765	-10868,7184	15	12	5	10,84	3,16	1164,6
16	7488	0,2160	0,0477	-22785,8265	16	12	6	11,18	3,82	1653,4
17Q	531	0,0290	0,0269	-1920,1158	17	13	6	11,31	4,69	492,7
17V	2048	0,2313	0,0475	-6242,0383	17	13	7	10,83	5,17	420,2
18Q	314	0,0432	0,0263	-1142,9315	18	14	7	11,89	5,11	191,0
18V	3462	0,1998	0,0387	-11259,0149	18	14	8	11,92	5,08	670,4
19	719	0,0345	0,0255	-2637,1936	19	15	8	12,23	5,77	531,7
20	3692	0,1836	0,0289	-13079,5405	20	15	9	13,24	5,76	581,8
21	287	0,0175	0,0246	-1063,5451	21	16	9	13,21	6,79	403,5
22	329	0,0204	0,0234	-1235,4073	22	17	9	14,22	6,78	377,3
23Q	196	0,0477	0,0222	-746,7074	23	18	9	15,45	6,55	91,1
23V	1710	0,1471	0,0002	-14278,9601	23	18	10	15,10	6,9	2,7
24	1738	0,1257	0,0002	-14782,8591	24	18	11	15,51	7,49	2,8
25	214	0,0177	0,0220	-816,3139	25	19	11	15,49	8,51	266,6
26Q	177	0,0175	0,0221	-674,9031	26	20	11	15,49	8,51	223,7
26V	229	0,1309	0,0000	-2526,9100	26	20	12	16,51	8,49	0,0
27	247	0,0215	0,0201	-964,7154	27	21	12	17,36	8,64	230,9
28	1475	0,1314	0,0001	-13227,5436	28	21	13	18,37	8,63	1,4
29	99	0,0102	0,0193	-390,7585	29	22	13	18,34	9,66	188,3
30	80	0,0083	0,0184	-319,5241	30	23	13	19,34	9,66	177,8
31	101	0,0106	0,0175	-408,3620	31	24	13	20,34	9,66	167,9
32	951	0,1004	0,0001	-8958,4990	32	24	14	21,37	9,63	0,8
33	368	0,0432	0,0164	-1513,5760	33	25	14	21,49	10,51	139,4
34	177	0,0217	0,0161	-731,3137	34	26	14	22,08	10,92	130,9
35	120	0,0150	0,0154	-501,1513	35	27	14	23,08	10,92	122,5
36	908	0,1156	0,0001	-8734,8132	36	27	15	24,11	10,89	0,5
37	13	0,0019	0,0147	-54,8218	37	28	15	24,09	11,91	102,4
38Q	51	0,0089	0,0140	-217,8572	38	29	15	25,16	11,84	80,0
38V	111	0,0924	0,0000	-1363,4735	38	29	16	24,70	12,3	0,0
39	593	0,0876	0,0130	-2574,2747	39	30	16	25,93	12,07	88,2
40	58	0,0094	0,0000	-776,6417	40	30	17	26,95	12,05	0,0
41	47	0,0077	0,0128	-204,7921	41	31	17	26,95	13,05	78,4

Figure 5.48.1: Performance du modèle beta géométrique 2





Le modèle béta géométrique 2 améliore la vraisemblance par rapport au modèle précédent (-499904 versus –554500 pour le modèle modèle béta géométrique 1.

On constate dans les figures 5.48.1 et 5.48.2 une meilleure représentation des observées par rapport aux résultats des modèles précédents.sur les premières questions.

En revanche, la probabilité d'arrêt aux questions suivantes est mal estimée

Par conséquent, même si le modèle Béta Géométrique 2 donne de meilleurs résultats, il n'est toujours pas satisfaisant.

Dans le troisième modèle que nous allons tester (modèle béta géométrique 3), nous considérons que les modèles des questions de qualification et des questions virales ne sont pas les mêmes, comme le laisse supposer la figure 5.38.
Pour les questions de qualification, nous avons simplifié le modèle et considéré que la probabilité d'arrêt est uniquement liée au nombre de questions de qualification auxquelles le joueur a déjà répondu.
Pour les questions virales, nous constatons une décroissance de la probabilité d'arrêt à une question virale avec un certain effet de seuil (figure 5.38) en fin de questionnaire.
Nous proposons donc le modèle suivant :

Modèle Béta géométrique 3 :

Question de qualification:
Proba (arrêt) = $(1-q_2)^{n2} q_2$

Question virale:
Proba (arrêt) = $\lambda_0 \ e^{-\gamma n1}$

avec

- n_1 : nombre de questions virales auxquelles l'individu a déjà répondu lorsqu'on lui pose la question N
- n_2 : nombre de questions de qualification auxquelles l'individu a déjà répondu par l'individu lorsqu'on lui pose la question N
- q_2 la probabilité d'arrêt du joueur à une question de qualification (q_2 suit une loi Béta)
- λ_0 et γ sont les paramètres du modèle utilisé pour la probabilité d'arrêt à une question virale.

Les formules et résultats du modèle sont présentés ci-dessous.

Tableau 5.62 : Formules modèle béta géométrique 3

lambda	0,35
gamma	0,06
alpha	0,00128
beta	0,000095

LL =SOMME(J9:J5

Q	Non réponses	Proba réelle	type estion	P(T=x)	LL	n2	n1	E(n_{1x})
1	13111	=F9/A9	1	=((1-LOI.BETA(F9;F$4;F$5;0;227879))^N9)*(LOI.BETA(F9;F$4;F$5;0;227879))	=F9*LN(I9)	0	0	=A9*I9
2	13146	=F10/A10	1	=((1-LOI.BETA(F10;F$4;F$5;0;227879))^N10)*(LOI.BETA(F10;F$4;F$5;0;227879))	=F10*LN(I10)	1	0	=A10*I10
3	14729	=F11/A11	1	=((1-LOI.BETA(F11;F$4;F$5;0;227879))^N11)*(LOI.BETA(F11;F$4;F$5;0;227879))	=F11*LN(I11)	2	0	=A11*I11
4	59518	=F12/A12	2	=F$2*EXP(-F$3*M12)	=F12*LN(I16)	3	0	=A12*I12
5	5216	=F13/A13	1	=((1-LOI.BETA(F13;F$4;F$5;0;227879))^N13)*(LOI.BETA(F13;F$4;F$5;0;227879))	=F13*LN(I13)	3	1	=A13*I13
6	1896	=F14/A14	1	=((1-LOI.BETA(F14;F$4;F$5;0;227879))^N14)*(LOI.BETA(F14;F$4;F$5;0;227879))	=F14*LN(I14)	4	1	=A14*I14
7	4546	=F15/A15	1	=((1-LOI.BETA(F15;F$4;F$5;0;227879))^N15)*(LOI.BETA(F15;F$4;F$5;0;227879))	=F15*LN(I15)	5	1	=A15*I15
8	38349	=F16/A16	2	=F$2*EXP(-F$3*M16)	=F16*LN(I16)	6	1	=A16*I16
9	2431	=F17/A17	1	=((1-LOI.BETA(F17;F$4;F$5;0;227879))^N17)*(LOI.BETA(F17;F$4;F$5;0;227879))	=F17*LN(I17)	6	2	=A17*I17
10	2810	=F18/A18	1	=((1-LOI.BETA(F18;F$4;F$5;0;227879))^N18)*(LOI.BETA(F18;F$4;F$5;0;227879))	=F18*LN(I18)	7	2	=A18*I18
11	21460	=F19/A19	2	=F$2*EXP(-F$3*M19)	=F19*LN(I19)	8	2	=A19*I19
12	1305	=F20/A20	1	=((1-LOI.BETA(F20;F$4;F$5;0;227879))^N20)*(LOI.BETA(F20;F$4;F$5;0;227879))	=F20*LN(I20)	8	3	=A20*I20
13Q	1679	=F21/A21	1	=((1-LOI.BETA(F21;F$4;F$5;0;227879))^N21)*(LOI.BETA(F21;F$4;F$5;0;227879))	=F21*LN(I21)	9,11	2,88414	=A21*I21
13V	5457	=F22/A22	2	=F$2*EXP(-F$3*M22)	=F22*LN(I22)	9	3	=A22*I22
14	1498	=F23/A23	1	=((1-LOI.BETA(F23;F$4;F$5;0;227879))^N23)*(LOI.BETA(F23;F$4;F$5;0;227879))	=F23*LN(I23)	9,55	3,44	=A23*I23
15Q	=E24	=F24/A24	1	=((1-LOI.BETA(F24;F$4;F$5;0;227879))^N24)*(LOI.BETA(F24;F$4;F$5;0;227879))	=F24*LN(I24)	10,3	3,62	=A24*I24
15V	=E25	=F25/A25	2	=F$2*EXP(-F$3*M25)	=F25*LN(I25)	10,8	3,16	=A25*I25
16	=E26	=F26/A26	2	=F$2*EXP(-F$3*M26)	=F26*LN(I26)	11,1	3,82	=A26*I26
17Q	=E27	=F27/A27	1	=((1-LOI.BETA(F27;F$4;F$5;0;227879))^N27)*(LOI.BETA(F27;F$4;F$5;0;227879))	=F27*LN(I27)	11,3	4,69	=A27*I27
17V	=E28	=F28/A28	2	=F$2*EXP(-F$3*M28)	=F28*LN(I28)	10,8	5,17	=A28*I28
18Q	=E29	=F29/A29	1	=((1-LOI.BETA(F29;F$4;F$5;0;227879))^N29)*(LOI.BETA(F29;F$4;F$5;0;227879))	=F29*LN(I29)	11,8	5,11	=A29*I29
18V	=E30	=F30/A30	2	=F$2*EXP(-F$3*M30)	=F30*LN(I30)	11,9	5,08	=A30*I30
19	=E31	=F31/A31	1	=((1-LOI.BETA(F31;F$4;F$5;0;227879))^N31)*(LOI.BETA(F31;F$4;F$5;0;227879))	=F31*LN(I31)	12,2	5,77	=A31*I31
20	=E32	=F32/A32	2	=F$2*EXP(-F$3*M32)	=F32*LN(I32)	13,2	5,76	=A32*I32
21	=E33	=F33/A33	1	=((1-LOI.BETA(F33;F$4;F$5;0;227879))^N33)*(LOI.BETA(F33;F$4;F$5;0;227879))	=F33*LN(I33)	13,2	6,79	=A33*I33
22	=E34	=F34/A34	1	=((1-LOI.BETA(F34;F$4;F$5;0;227879))^N34)*(LOI.BETA(F34;F$4;F$5;0;227879))	=F34*LN(I34)	14,2	6,78	=A34*I34
23Q	=E35	=F35/A35	1	=((1-LOI.BETA(F35;F$4;F$5;0;227879))^N35)*(LOI.BETA(F35;F$4;F$5;0;227879))	=F35*LN(I35)	15,4	6,55	=A35*I35
23V	=E36	=F36/A36	2	=F$2*EXP(-F$3*M36)	=F36*LN(I36)	15,1	6,9	=A36*I36
24	=E37	=F37/A37	2	=F$2*EXP(-F$3*M37)	=F37*LN(I37)	15,5	7,49	=A37*I37
25	=E38	=F38/A38	1	=((1-LOI.BETA(F38;F$4;F$5;0;227879))^N38)*(LOI.BETA(F38;F$4;F$5;0;227879))	=F38*LN(I38)	15,4	8,51	=A38*I38
26Q	=E39	=F39/A39	1	=((1-LOI.BETA(F39;F$4;F$5;0;227879))^N39)*(LOI.BETA(F39;F$4;F$5;0;227879))	=F39*LN(I39)	15,4	8,51	=A39*I39
26V	=E40	=F40/A40	2	=F$2*EXP(-F$3*M40)	=F40*LN(I40)	16,5	8,49	=A40*I40
27	=E41	=F41/A41	1	=((1-LOI.BETA(F41;F$4;F$5;0;227879))^N41)*(LOI.BETA(F41;F$4;F$5;0;227879))	=F41*LN(I41)	17,3	8,64	=A41*I41
28	=E42	=F42/A42	2	=F$2*EXP(-F$3*M42)	=F42*LN(I42)	18,3	8,63	=A42*I42
29	=E43	=F43/A43	1	=((1-LOI.BETA(F43;F$4;F$5;0;227879))^N43)*(LOI.BETA(F43;F$4;F$5;0;227879))	=F43*LN(I43)	18,3	9,66	=A43*I43
30	=E44	=F44/A44	1	=((1-LOI.BETA(F44;F$4;F$5;0;227879))^N44)*(LOI.BETA(F44;F$4;F$5;0;227879))	=F44*LN(I44)	19,3	9,66	=A44*I44
31	=E45	=F45/A45	1	=((1-LOI.BETA(F45;F$4;F$5;0;227879))^N45)*(LOI.BETA(F45;F$4;F$5;0;227879))	=F45*LN(I45)	20,3	9,66	=A45*I45
32	=E46	=F46/A46	2	=F$2*EXP(-F$3*M46)	=F46*LN(I46)	21,3	9,63	=A46*I46
33	=E47	=F47/A47	1	=((1-LOI.BETA(F47;F$4;F$5;0;227879))^N47)*(LOI.BETA(F47;F$4;F$5;0;227879))	=F47*LN(I47)	21,4	10,51	=A47*I47
34	=E48	=F48/A48	1	=((1-LOI.BETA(F48;F$4;F$5;0;227879))^N48)*(LOI.BETA(F48;F$4;F$5;0;227879))	=F48*LN(I48)	22,0	10,92	=A48*I48
35	=E49	=F49/A49	1	=((1-LOI.BETA(F49;F$4;F$5;0;227879))^N49)*(LOI.BETA(F49;F$4;F$5;0;227879))	=F49*LN(I49)	23,0	10,92	=A49*I49
36	=E50	=F50/A50	2	=F$2*EXP(-F$3*M50)	=F50*LN(I50)	24,1	10,89	=A50*I50
37	=E51	=F51/A51	1	=((1-LOI.BETA(F51;F$4;F$5;0;227879))^N51)*(LOI.BETA(F51;F$4;F$5;0;227879))	=F51*LN(I51)	24,0	11,91	=A51*I51
38Q	=E52	=F52/A52	1	=((1-LOI.BETA(F52;F$4;F$5;0;227879))^N52)*(LOI.BETA(F52;F$4;F$5;0;227879))	=F52*LN(I52)	25,1	11,84	=A52*I52
38V	=E53	=F53/A53	2	=F$2*EXP(-F$3*M53)	=F53*LN(I53)	24,7	12,3	=A53*I53
39	=E54	=F54/A54	1	=((1-LOI.BETA(F54;F$4;F$5;0;227879))^N54)*(LOI.BETA(F54;F$4;F$5;0;227879))	=F54*LN(I54)	25,9	12,07	=A54*I54
40	=E55	=F55/A55	2	=F$2*EXP(-F$3*M55)	=F55*LN(I55)	26,9	12,05	=A55*I55
41	=E56	=F56/A56	1	=((1-LOI.BETA(F56;F$4;F$5;0;227879))^N56)*(LOI.BETA(F56;F$4;F$5;0;227879))	=F56*LN(I56)	26,9	13,05	=A56*I56

Tableau 5.63: Résultats modèle béta géométrique 3

lambda0	0,35
gamma	0,06
alpha	0,00128
beta	0,00010

LL -394034,5798

Q	Non réponses	Proba réelle	type question	P(T=x)	LL	question	question virale	n2	n1	E(n_{1x})
1	13111	0,0575	1	0,0688	-35083,8697	1	0	0,00	0	15688,1
2	13146	0,0612	1	0,0641	-36115,1675	2	0	1,00	0	13767,7
3	14729	0,0731	1	0,0597	-41512,6991	3	0	2,00	0	12036,7
4	59518	0,3185	2	0,3296	-69625,4732	3	1	3,00	0	61603,2
5	5216	0,0409	1	0,0555	-15078,6781	4	1	3,00	1	7073,0
6	1896	0,0155	1	0,0517	-5617,8768	5	1	4,00	1	6311,1
7	4546	0,0378	1	0,0481	-13790,0193	6	1	5,00	1	5790,6
8	38349	0,3314	2	0,3104	-44861,5087	6	2	6,00	1	35921,1
9	2431	0,0314	1	0,0448	-7548,5421	7	2	6,00	2	3467,6
10	2810	0,0375	1	0,0417	-8925,0997	8	2	7,00	2	3128,2
11	21460	0,2975	2	0,2923	-26391,9828	8	3	8,00	2	21085,9
12	1305	0,0258	1	0,0389	-4238,3486	9	3	8,00	3	1968,9
13Q	1679	0,0684	1	0,0359	-5586,0558	10	3	9,12	2,88	881,4
13V	5457	0,2199	2	0,2753	-7038,5593	10	4	9,00	3	6831,0
14	1498	0,0355	1	0,0348	-5030,7451	11	4	9,56	3,44	1469,2
15Q	1833	0,0719	1	0,0328	-6262,5375	12	4	10,38	3,62	837,1
15V	4228	0,2777	2	0,2593	-5707,0479	12	5	10,84	3,16	3948,2
16	7488	0,2160	2	0,2442	-10556,7481	12	6	11,18	3,82	8465,2
17Q	531	0,0290	1	0,0307	-1849,7215	13	6	11,31	4,69	562,6
17V	2048	0,2313	2	0,2300	-3010,1957	13	7	10,83	5,17	2036,1
18Q	314	0,0432	1	0,0295	-1106,6283	14	7	11,89	5,11	214,4
18V	3462	0,1998	2	0,2166	-5296,2442	14	8	11,92	5,08	3752,6
19	719	0,0345	1	0,0288	-2551,7164	15	8	12,23	5,77	598,8
20	3692	0,1836	2	0,2040	-5869,6233	15	9	13,24	5,76	4100,7
21	287	0,0175	1	0,0268	-1038,5421	16	9	13,21	6,79	440,2
22	329	0,0204	1	0,0250	-1213,8902	17	9	14,22	6,78	402,8
23Q	196	0,0477	1	0,0229	-740,3142	18	9	15,45	6,55	94,1
23V	1710	0,1471	2	0,1921	-2821,1958	18	10	15,10	6,9	2232,8
24	1738	0,1257	2	0,1809	-2971,6709	18	11	15,51	7,49	2502,0
25	214	0,0177	1	0,0228	-808,8850	19	11	15,49	8,51	276,0
26Q	177	0,0175	1	0,0228	-669,0251	20	11	15,49	8,51	231,2
26V	229	0,1309	2	0,1704	-405,2893	20	12	16,51	8,49	298,0
27	247	0,0215	1	0,0200	-966,3764	21	12	17,36	8,64	229,4
28	1475	0,1314	2	0,1604	-2698,9876	21	13	18,37	8,63	1801,1
29	99	0,0102	1	0,0187	-394,2015	22	13	18,34	9,66	181,9
30	80	0,0083	1	0,0174	-324,2184	23	13	19,34	9,66	167,7
31	101	0,0106	1	0,0162	-416,5104	24	13	20,34	9,66	154,9
32	951	0,1004	2	0,1511	-1797,2208	24	14	21,37	9,63	1431,1
33	368	0,0432	1	0,0149	-1547,9493	25	14	21,49	10,51	126,9
34	177	0,0217	1	0,0143	-751,7965	26	14	22,08	10,92	116,6
35	120	0,0150	1	0,0133	-518,1493	27	14	23,08	10,92	106,3
36	908	0,1156	2	0,1423	-1770,4385	27	15	24,11	10,89	1117,8
37	13	0,0019	1	0,0124	-57,0306	28	15	24,09	11,91	86,4
38Q	51	0,0089	1	0,0115	-227,6825	29	15	25,16	11,84	66,0
38V	111	0,0924	2	0,1340	-223,0903	29	16	24,70	12,3	160,9
39	593	0,0876	1	0,0109	-2681,2146	30	16	25,93	12,07	73,6
40	58	0,0094	2	0,1262	-120,0497	30	17	26,95	12,05	779,8
41	47	0,0077	1	0,0101	-215,7624	31	17	26,95	13,05	62,1

On constate que le modèle Béta géométrique 3 est bien meilleur que les modèles précédents :

- la vraisemblance est nettement améliorée (-390034 versus - 499904 pour le modèle Béta géométrique 2 et une vraisemblance inférieure pour les autres modèles testés).

- le nombre de paramètres est moins élevé que le modèle Béta Géométrique 2 (4 paramètres versus 5).

Ce modèle est également intéressant car il reflète sur les questions virales la baisse de la probabilité d'arrêt au fur et à mesure du nombre de questions virales auxquelles le joueur a répondu. L'hétérogénéité des joueurs (qui explique ce phénomène) est intégrée directement dans la formule utilisée dans le cas des questions virales, alors que cette hétérogénéité est représentée pour les questions de qualification par le fait que q_2 (probabilité d'arrêt à une question de qualification) suit une loi béta. D'autre part, on peut penser que plus les joueurs répondent aux questions virales, moins il leur reste d'adresses e-mail à fournir car leur stock d'adresses e-mail est limité (amis, collègues,…). Ce phénomène est dans notre modèle Béta Géométrique 3 moins important que l'effet d'hétérogénéité, puisqu'au global la probabilité d'arrêt aux questions virales diminue en fonction du nombre de questions virales auxquelles le joueur à répondu. Enfin, le dernier avantage de ce modèle est qu'il ne nécessite pas de traitement particulier pour la première question, à la différence du modèle Béta Géométrique 2 que nous avons proposé précédemment.

C'est par conséquent le modèle Béta Géométrique 3 que nous allons retenir pour représenter la probabilité d'arrêt des joueurs aux différents questions du jeu.

Les paramètres de ce modèle sont les suivants :

Tableau 5.64: Paramètres du modèle Béta Géométrique 3

lambda0	0,35
gamma	0,06
alpha	0,0012
beta	0,0001

5.4.3 Modèle de transformation des adresses virales saisies en adresses virales joueurs

L'objectif de cette partie est de modéliser la loi de la transition du type *ViralQ4joue* du réseau de Petri de la figure 5.2, ce qui revient à calculer la loi de probabilité pour qu'une adresse virale fournie par un joueur se mette à jouer à son tour.

Pour définir cette transition dans le réseau de Petri, nous devons modéliser deux variables :

1. La probabilité qu'une adresse virale saisie joue (qui nous permettra par exemple sur la figure 5.2 d'affecter des probabilités aux transitions *ViralQ4joue* et *ViralQ4ne joue pas*, et donc de savoir combien d'adresses virales fournies ont ou n'ont pas joué)
2. La durée entre le moment auquel l'adresse virale est saisie et le moment où elle s'inscrit au jeu (ce qui nous permettra de paramétrer la loi de la transition *ViralQ4joue* en introduisant cette durée dans le modèle).

5.4.3.1 Loi de la probabilité qu'une adresse virale saisie joue:

L'observation des données concernant le taux de transformation d'adresses virales en joueurs

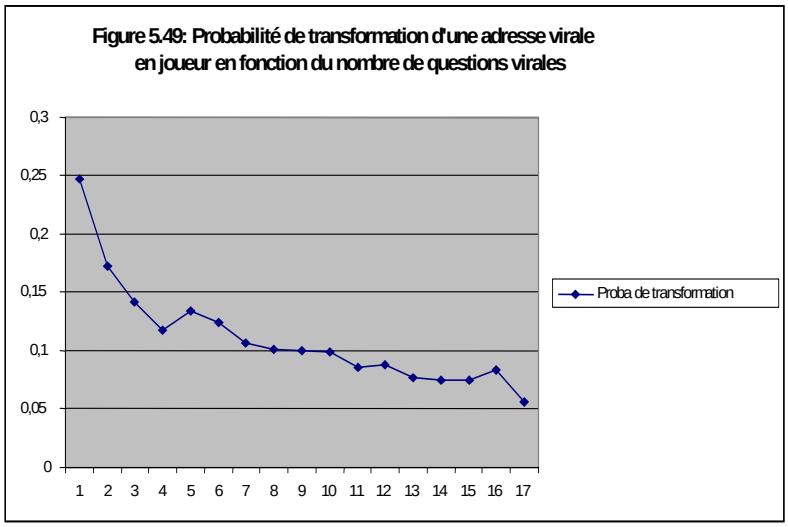
Tableau 5.62: Probabilité de transformation par question

	question	Nbre réponse	nbre viral joueur	proba viral joue
1	Q4	127375	31498	0,2473
2	Q8	77368	13347	0,1725
3	Q11	50667	7207	0,1422
4	Q13V	19130	2258	0,118
5	Q15V	10999	1481	0,1346
6	Q16	26631	3302	0,124
7	Q17V	6806	729	0,1071
8	Q18V	13865	1405	0,1013
9	Q20	16413	1649	0,1005
10	Q23V	9914	979	0,0987
11	Q24	12093	1041	0,0861
12	Q26V	1520	133	0,0875
13	Q28	9751	748	0,0767
14	Q32	8520	639	0,075
15	Q36	6947	520	0,0749
16	Q38V	1090	91	0,0835
17	Q40	6121	345	0,0564

est présentée dans le tableau 5.62.

On constate tout d'abord que le taux de transformation décroît au fur et à mesure des questions virales posées. On peut aussi toutefois noter que même en Q40 le taux de transformation reste quand même non négligeable (5,64%).

Le pourcentage d'adresses virales transformées en joueurs suit la courbe de la figure 5.50, sur laquelle on constate une forte baisse sur les premières questions virales puis une décroissance plus atténuée sur les autres questions virales (la question 1 correspond à la première question virale du questionnaire, soit Q4 ; la question 17 correspond à la dernière question virale du questionnaire, soit Q40).



Nous avons testé plusieurs modèles qui sont présentés ci-dessous, en commençant par un modèle exponentiel pour représenter par question la probabilité de transformation d'une adresse virale saisie.

Tableau 5.63 : Résultats du modèle exponentiel :

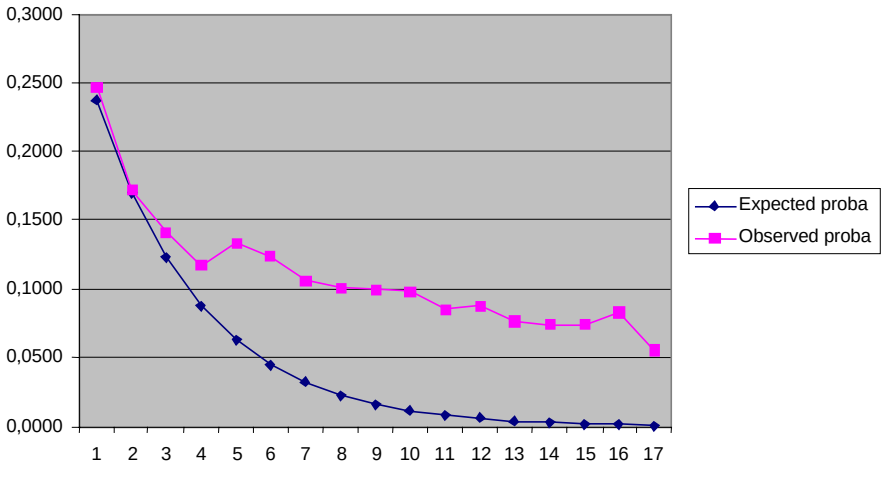
lambda 0,330459003

LL -141971,1733 Loi Exponentielle

Nbre réponse	Question	Nbre transformés	P(X=x)	observé	LL	Expected
127375	1	31498	0,2375	0,2473	-45285,6723	30247,2060
77368	2	13347	0,1706	0,1725	-23600,0410	13202,1881
50667	3	7207	0,1226	0,1422	-15124,9683	6212,8836
19130	4	2258	0,0881	0,118	-5484,9274	1685,6464
10999	5	1481	0,0633	0,1346	-4086,9197	696,4472
26631	6	3302	0,0455	0,124	-10203,2681	1211,7302
6806	7	729	0,0327	0,1071	-2493,5341	222,5327
13865	8	1405	0,0235	0,1013	-5270,0773	325,7657
16413	9	1649	0,0169	0,1005	-6730,2347	277,1130
9914	10	979	0,0121	0,0987	-4319,2136	120,2822
12093	11	1041	0,0087	0,0861	-4936,7569	105,4314
1520	12	133	0,0063	0,0875	-674,6798	9,5228
9751	13	748	0,0045	0,0767	-4041,6233	43,8987
8520	14	639	0,0032	0,075	-3663,8335	27,5630
6947	15	520	0,0023	0,0749	-3153,3620	16,1498
1090	16	91	0,0017	0,0835	-581,9101	1,8209
6121	17	345	0,0012	0,0564	-2320,1511	7,3478

Performance du modèle :

Figure 5.50: Probabilité qu'une adresse virale joue par question virale



Le modèle prévoit correctement sur les trois premières questions virales posées. Il est beaucoup moins performant sur les dernières questions virales posées dans le questionnaire.

Afin d'améliorer l'estimation, nous avons testé plusieurs autres modèles:

- un modèle exponentiel gamma (en faisant l'hypothèse que les individus ne sont pas les mêmes d'une question à l'autre)
- un modèle exponentiel gamma avec 1 covarié (le numéro de la question virale). Ce choix se fonde sur l'hypothèse selon laquelle le taux de transformation en joueurs est lié au nombre de questions virales auxquelles l'individu a déjà répondu. En effet, on peut penser que la « qualité »-en terme de joueurs potentiels- des adresses e-mail fournies s'érode probablement. Aux premières questions virales, l'internaute fournit des contacts proches qui vont sur ses conseils jouer lorsqu'ils vont recevoir l'e-card virale tandis qu'à la fin du questionnaire, les adresses virales fournies sont celles de personnes plus lointaines).
- Un modèle de Weibull
- Un modèle Weibull Gamma
- Un modèle Weibull Gamma + 1 Covarié (numéro de la question virale)

La comparaison des modèles s'effectue en appliquant le test du ratio de vraisemblance, présenté dans le tableau ci-dessous :

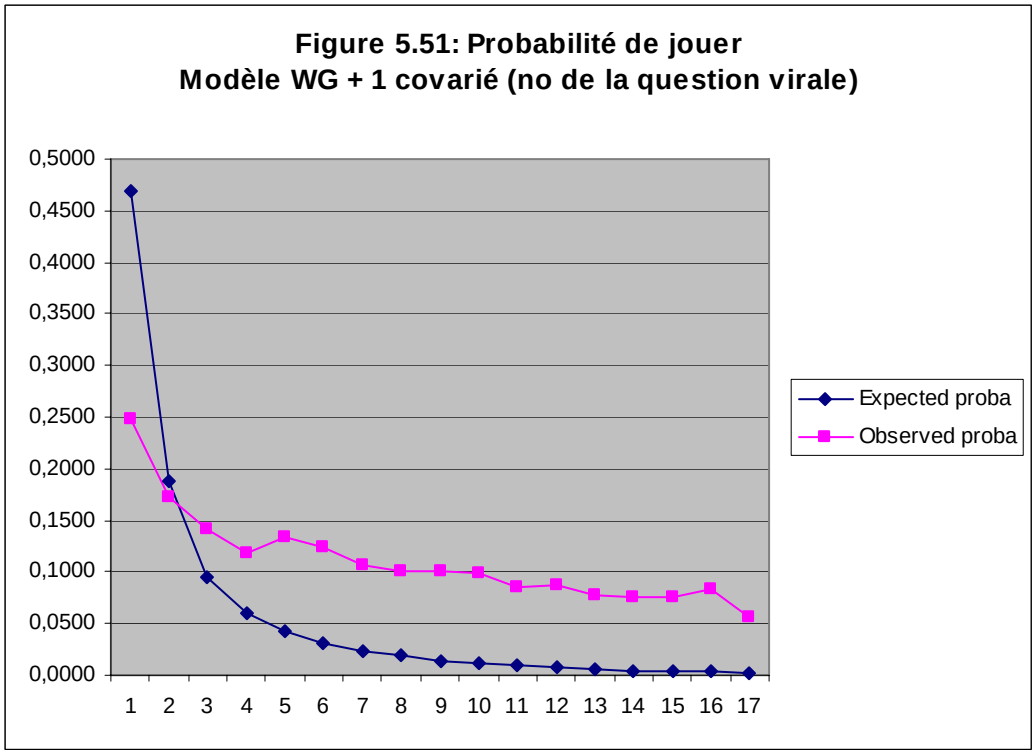
Tableau 5.64: Comparaison des modèles

	LL	Nbre de paramètres	LR	Khi 2 (0,05;1)	
Exponentiel	-141971	1			
Exponentiel Gamma	-125030	2	33881	3,841	*
EG+ 1 covarié (numéro)	-123770	3	2520	3,841	**
Weibull	-141357	2	1228	3,841	*
Weibull Gamma	-124190	3	1681	3,841	***
Weibull Gamma + 1 covarié (numéro)	-123669	4	202	3,841	***
Weibull Gamma + 1 covarié (numéro)	-123669	4	1041	3,841	****

* Comparé au modèle exponentiel
** Comparé au modèle exponentiel gamma
*** Comparé au modèle exponentiel gamma + 1 covarié
**** Comparé au modèle weibull gamma

Ce tableau montre que le modèle Weibull Gamma + 1 covarié est meilleur que les modèles exponentiel, exponentiel gamma et weibull gamma + 1 covarié (qui est lui même meilleur que le modèle simple weibull Gamma).

Le modèle WG + 1 covarié, même s'il est meilleur que les autres modèles ci-dessus cités, n'est pas satisfaisant pour représenter le taux de transformation des adresses virales, comme le montre la figure 5.51.



On constate sur la figure 5.51 qu'un modèle exponentiel ou weibull ne représente pas le phénomène car ces modèles décroissent jusqu'à la valeur 0, alors que dans notre cas, la probabilité de joueur décroît fortement au début puis baisse plus progressivement avec un effet de « seuil » (il reste 5% d'adresses transformées à la 17ème question virale posée, chiffre élevé au regard du nombre de questions virales posées).

Nous avons donc appliqué un modèle de type exponentiel mais intégrant un effet de seuil.

La littérature en marketing comporte plusieurs modèles de ce type, et notamment le modèle « Log réciproque » utilisé par Bemmaor (1984) pour modéliser les effets de seuil lors de campagnes publicitaires ou encore par Brown et Tucker (1961).

Dans le cas où on cherche à modéliser des ventes en fonction d'investissements publicitaires, le modèle « Log réciproque » a la forme suivante :

$$Q=e^{a-\frac{b}{X}}$$

avec Q : ventes et X : investissements publicitaires.

Pour a>b et x>0, ce modèle décroît lorsque X augmente, ce qui correspond au phénomène que nous étudions. Nous avons cependant adapté ce modèle puisque nous mesurons dans notre cas une probabilité qui doit être comprise entre 0 et 1.

Par conséquent, nous avons établi le modèle suivant :

Probabilité de jouer = 1-q,

q suivant une loi exponentielle de paramètre (a-(b/n))

avec comme contraintes sur les paramètres :

a>0, b>0.

Probabilité de jouer :

$p=1-e^{a-\frac{b}{n}}$ avec n : numéro de la question posée.

Les formules sous excel et les résultats sont les suivants :

Tableau 5.64 : Formules du modèle taux de transformation

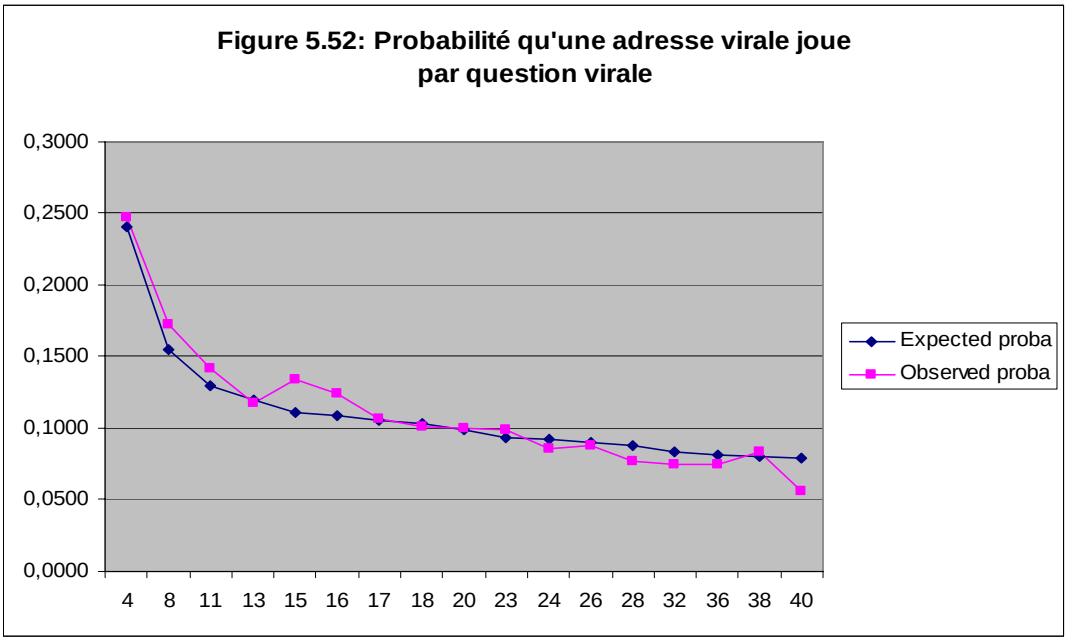
	a	0,09				
	b	0,86				
	LL	=SOMME(H6:H22)				
Nbre réponse	Question	Nbre transformés	P(X=x)	observé	LL	Expected
127375	4	31498	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D6;E$2;FAUX))	0,2473	=E6*LN(F6)	=C6*F6
77368	8	13347	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D7;E$2;FAUX))	0,1725	=E7*LN(F7)	=C7*F7
50667	11	7207	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D8;E$2;FAUX))	0,1422	=E8*LN(F8)	=C8*F8
19130	13	2258	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D9;E$2;FAUX))	0,118	=E9*LN(F9)	=C9*F9
10999	15	1481	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D10;E$2;FAUX))	0,1346	=E10*LN(F10)	=C10*F10
26631	16	3302	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D11;E$2;FAUX))	0,124	=E11*LN(F11)	=C11*F11
6806	17	729	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D12;E$2;FAUX))	0,1071	=E12*LN(F12)	=C12*F12
13865	18	1405	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D13;E$2;FAUX))	0,1013	=E13*LN(F13)	=C13*F13
16413	20	1649	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D14;E$2;FAUX))	0,1005	=E14*LN(F14)	=C14*F14
9914	23	979	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D15;E$2;FAUX))	0,0987	=E15*LN(F15)	=C15*F15
12093	24	1041	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D16;E$2;FAUX))	0,0861	=E16*LN(F16)	=C16*F16
1520	26	133	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D17;E$2;FAUX))	0,0875	=E17*LN(F17)	=C17*F17
9751	28	748	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D18;E$2;FAUX))	0,0767	=E18*LN(F18)	=C18*F18
8520	32	639	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D19;E$2;FAUX))	0,075	=E19*LN(F19)	=C19*F19
6947	36	520	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D20;E$2;FAUX))	0,0749	=E20*LN(F20)	=C20*F20
1090	38	91	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D21;E$2;FAUX))	0,0835	=E21*LN(F21)	=C21*F21
6121	40	345	=1-(EXP(E$1)*LOI.EXPONENTIELLE(1/D22;E$2;FAUX))	0,0564	=E22*LN(F22)	=C22*F22

Tableau 5.65 : Résultats du modèle taux de transformation

a	0,09
b	0,86
LL	-119406,5268

Nbre réponse	Question	Nbre transform	P(X=x)	observé	LL	Expected
127375	4	31498	0,2411	0,2473	-44813,4597	30704,0830
77368	8	13347	0,1549	0,1725	-24890,1847	11985,7792
50667	11	7207	0,1298	0,1422	-14716,2966	6575,3641
19130	13	2258	0,1192	0,118	-4801,7907	2281,1729
10999	15	1481	0,1114	0,1346	-3249,6755	1225,7590
26631	16	3302	0,1083	0,124	-7341,2825	2882,8873
6806	17	729	0,1054	0,1071	-1640,0433	717,5505
13865	18	1405	0,1029	0,1013	-3194,8108	1426,8668
16413	20	1649	0,0986	0,1005	-3819,9599	1618,5692
9914	23	979	0,0935	0,0987	-2319,5542	927,4078
12093	24	1041	0,0921	0,0861	-2482,2994	1114,1516
1520	26	133	0,0896	0,0875	-320,8106	136,2316
9751	28	748	0,0875	0,0767	-1822,4493	852,9454
8520	32	639	0,0840	0,075	-1583,0496	715,3598
6947	36	520	0,0812	0,0749	-1305,4899	564,2558
1090	38	91	0,0801	0,0835	-229,7651	87,2731
6121	40	345	0,0790	0,0564	-875,6049	483,7149

Nous retiendrons donc ce modèle pour mesurer le taux de transformation des adresses virales en joueurs puisqu'il donne de bonnes prévisions, comme le confirme la figure ci-dessous :

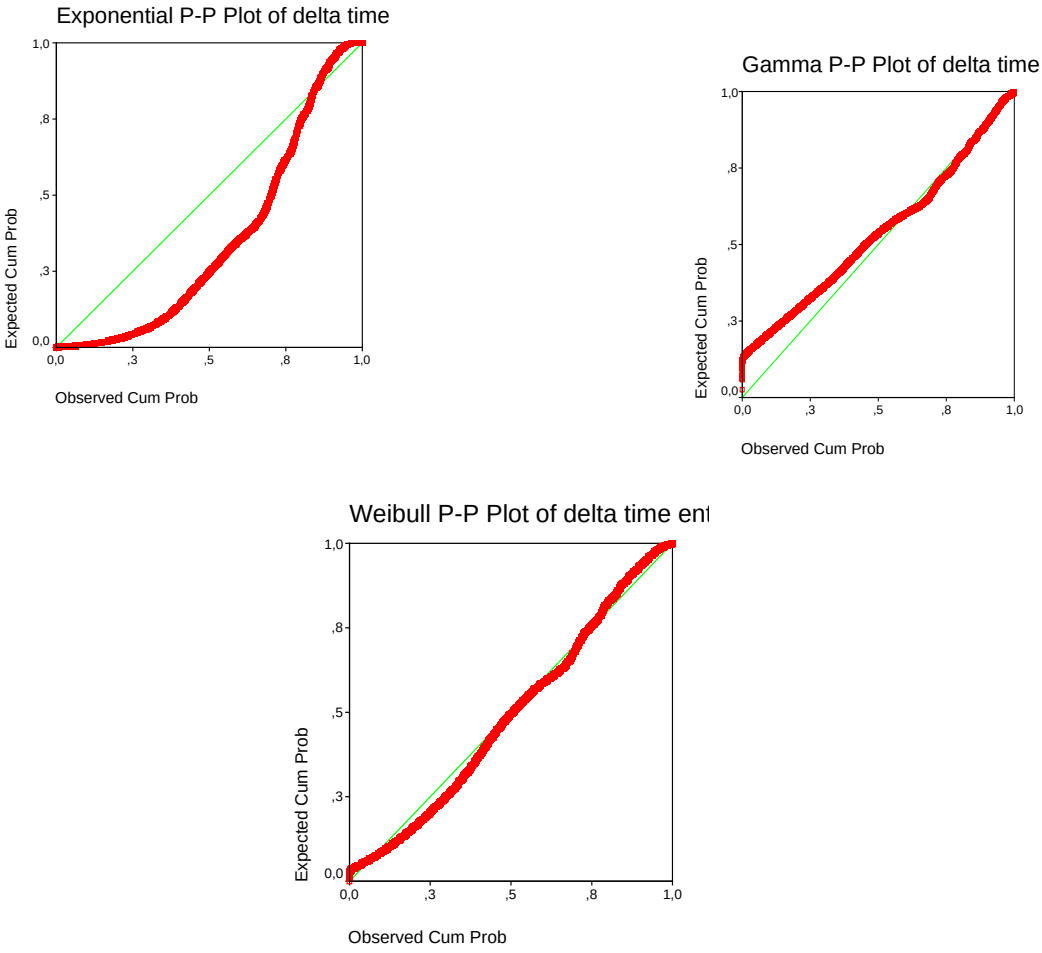


5.4.3.2 Durée entre saisie adresse virale et inscription au jeu de l'adresse virale :

L'objectif ici est de modéliser la loi de la durée entre la saisie de l'adresse virale et le moment auquel l'adresse virale s'inscrit au jeu (en heure). En moyenne, cette durée est de 2 jours avec un écart-type de 3,86.

L'analyse des PP Plot de cette variable (en heure) permet de sélectionner les lois les plus proches des données observées, qui sont représentées dans la figure 5.52.

Figure 5.66 : PP plot de la variable : durée entre adresse saisie et adresse joueuse



On conclut tout d'abord au rejet de la loi exponentielle. Les loi Weibull et Gamma représentent bien les données observées. Ces deux lois ont donc été étudiées et leurs paramètres ont été calculés sur Excel sur la base des 54507 données disponibles (données exprimées en secondes), et grâce à la méthode du maximum de vraisemblance.
La loi Gamma donne une vraisemblance de –72825 et la loi de Weibull –72652.
Nous retiendrons donc cette loi de Weibull pour représenter la loi de la durée entre le moment où une adresse virale est fournie dans le jeu et le moment auquel cette adresse virale joue effectivement.

Les paramètres obtenus pour la loi de Weibull sont les suivants :

Tableau 5.67: Paramètres Pétri de la loi de Weibull
Durée entre adresse virale fournie et adresse virale joueur

alpha Pétri	0,468804
béta Pétri	0,005075

Rappelons que l'équation de la fonction de densité de la loi de Weibull dans les réseaux de Petri est:

$$f(x;\alpha;\beta)=\alpha\beta x^{\beta-1} e^{-\alpha * x^{\beta}}$$

CONCLUSION :

L'objectif de ce chapitre était de modèliser le jeu promotionnel viral on line « Le ciel est à tout le monde » effectué par la société Millemercis pour le compte de son client Anyway. Nous avons estimé plusieurs modèles :

1. les inscriptions des joueurs : chaque source de recrutement a été modélisée, ce qui nous permet d'estimer les joueurs provenant des campagnes e-mailing (annonce et relance), de l'affiliation, de l'autopromotion sur le site Anyway, ou d'autres sources non identifiées.
2. les réponses aux questions : pour chaque question, nous estimons par des modèles le nombre de non réponses, ce qui nous a permis de mieux comprendre la manière dont les joueurs traitent le questionnaire.
3. La loi de transformation des adresses virales en joueurs ainsi que la loi de la durée entre une adresse e-mail fournie de manière virale et le moment auquel cette adresse virale joue : ceci nous permet de mieux comprendre l'aspect viral du jeu et son efficacité.

Nous avons été particulièrement vigilants dans ce chapitre à choisir le bon modèle dans chacun des cas ci-dessus. En effet, ces modèles représentent les lois des différentes transitions du réseau de Petri que nous souhaitons développer pour représenter le jeu de manière globale et effectuer des simulations. Les lois que nous avons trouvées doivent par conséquent être particulièrement robustes. Une synthèse de ces lois sera d'ailleurs présentée dans le prochain chapitre.

Etant donné que nous avons désormais tous les éléments pour construire le réseau de Petri du jeu Anyway, nous allons maintenant consacrer le chapitre 6 à l'intégration des modèles dans le réseau de Petri et à développer des simulations visant à faire des recommandations de stratégie marketing dans le cadre de telles opérations.

Chapitre 6:

Modélisation, mesure de la performance et simulation d'un jeu promotionnel viral on line par les réseaux de Petri

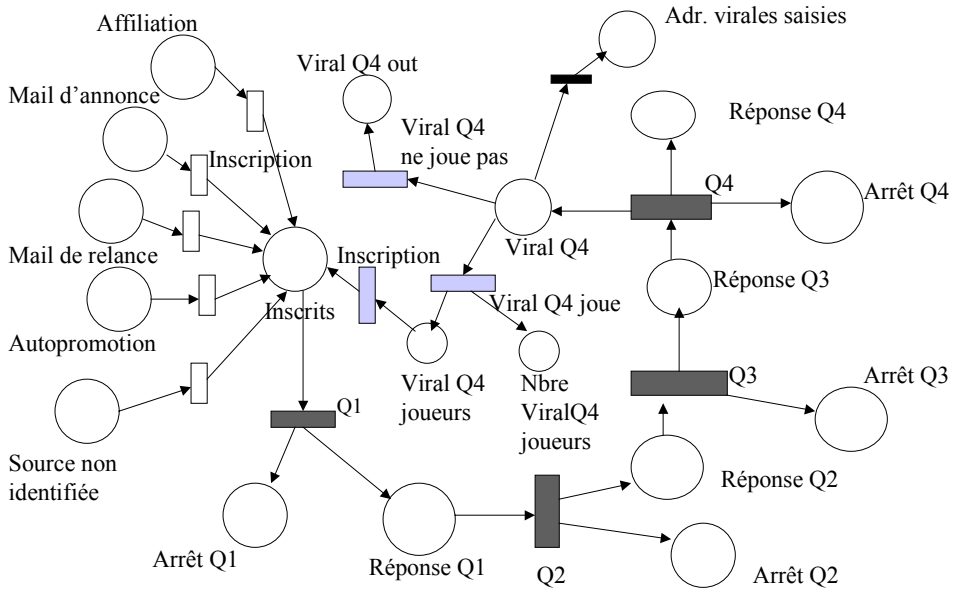
6.1 INTRODUCTION

Nous avons dans le chapitre précédent détaillé les différents modèles permettant de mieux décrire le comportement des joueurs du jeu promotionnel on line « le Ciel est à tout le monde ». Dans ce chapitre, notre objectif est d'intégrer l'ensemble de ces modèles dans un système global représentant le jeu dans son ensemble. Nous cherchons en effet à mesurer la performance globale du jeu et en particulier son efficacité en terme de marketing viral. De plus, nous visons à améliorer ce jeu et à donc effectuer plusieurs simulations de stratégies marketing différentes afin d'en mesurer l'impact et de choisir la meilleure de ces stratégies. Ce système global sera modélisé par un réseau de Petri que nous allons précisément décrire. Ce chapitre se décompose de la façon suivante : dans une première partie, nous décrirons les composantes du réseau de Petri utilisé pour modéliser le jeu « le Ciel est à tout le monde », et mesurerons la performance du modèle. Dans une deuxième partie, nous mènerons plusieurs simulations dans l'objectif d'améliorer les performances virales de ce jeu on line. Nous conclurons par des recommandations d'actions pour optimiser le mécanisme et le design de ce type d'opération de marketing viral sur Internet.

6.2 RESEAU DE PETRI DU JEU PROMOTIONNEL ON LINE

Comme nous l'avons décrit dans le chapitre précédent, l'opération « Le ciel est à tout le monde » peut être décomposée en un certain nombre d'actions séquentielles effectuées par l'internaute, ces actions correspondant à différents évènements intervenant sur le système. La figure 6.1 décrit le réseau de Petri représentant le jeu dans le cas simple des quatre premières questions. Rappelons le mécanisme du réseau que nous avons déjà présenté dans le précédent chapitre.

Figue 6.1: Réseau de Pétri de l'opération « Le ciel est à tout le monde » : exemple des 4 premières questions.



Les places affiliation, mail d'annonce, mail de relance, autopromotion et source non identifiée sont les places initiales du système. Ces places contiennent les internautes qui viennent sur le site du jeu promotionnel par chacune de ces sources de recrutement. Les transitions de couleur blanches représentent les inscriptions : la transition *inscription* s'effectue lorsqu'un joueur s'inscrit au jeu. Précisons que les couleurs des transitions de la figure 6.1 ont pour seul objectif de clarifier le schéma : les transitions blanches correspondent au phénomène des inscriptions ; les transitions grises foncées décrivent les réponses aux questions; les grises claires l'aspect viral du jeu : la transition noire est une transition immédiate (ne suivant pas de loi stochastique) qui se déclenche automatiquement dès qu'il y a un jeton dans la place *ViralQ4*. Nous avons montré dans le chapitre 5 que la loi des inscriptions est spécifique à chaque source de recrutement (affiliation, mail d'annonce, mail de relance, autopromotion, source non identifiée). Le joueur inscrit est alors placé dans la place *Inscrits*. Il est ensuite soumis à la question Q1, représentée par la transition verte *Q1* sur la figure 6.1. Si le joueur répond à la question Q1, il passe dans la place *réponseQ1*. En revanche, s'il s'arrête de jouer sans répondre à la question Q1, il passe dans la place *ArrêtQ1*. Dans le cas où le joueur a répondu à la question Q1, le jeu se poursuit : il est ensuite soumis à la question Q2 et ainsi de suite. Lorsque le joueur arrive à une question virale, comme la Q4 par exemple, il peut soit

répondre (place *RéponseQ4*) soit ne pas répondre (place *ArrêtQ4*), mais on va également pour cette question virale mesurer le nombre d'adresses virales fournies. En effet, chaque joueur répondant à la question Q4 passe par la transition Q4 qui crée un jeton dans la place *ViralQ4*, ce qui indique que ce joueur a fourni une adresse e-mail. L'adresse virale ainsi saisie reçoit un e-mail l'invitant à participer au jeu (ceci n'est pas représenté sur le réseau de Petri car c'est une action immédiate et systématique n'influençant pas les paramètres du modèle. De plus, le rajout de places et transitions pour décrire cet envoi d'e-mail compliquerait le graphique et diminuerait sa compréhension). Lorsqu'il reçoit l'e-mail d'invitation au jeu, l'internaute dont le joueur a fourni l'adresse virale a deux possibilités : soit il joue, soit il ne joue pas. S'il veut jouer, il passe par la transition gris clair *ViralQ4joue* puis dans la place *ViralQ4joueurs.* Il s'inscrit au jeu par la transition *Inscription* et rentre dans le mécanisme du jeu. Nous représentons donc bien la provenance des joueurs issus du marketing viral comme une variable endogène du système. On comptabilise le nombre d'adresses virales qui sont devenues joueurs (indicateur de transformation de viral en joueur) dans la place *NbreViralQ4joueurs.* S'il ne joue pas, il passe par la transition gris clair *ViralQ4nejouepas* puis dans la place *ViralQ4out* qui fournit donc le nombre d'adresses virales fournies n'ayant pas joué.

Les indicateurs de performance du jeu tirés du réseau de Petri de la figure 5.1 sont par conséquent :

- **le nombre d'adresses virales fournies par le jeu (nombre de jetons de la place *Adr.Virales saisies*),**
- **le nombre d'adresses virales ayant joué (nombre de jetons de la place *NbreViralQ4joueurs*),**
- **le nombre de non-réponses par question (nombre de jetons des places *ArrêtQ1, ArrêtQ2, ArrêtQ3, ArrêtQ4*).**

Chaque transition du réseau correspond à un événement dans le système qui suit une loi stochastique (exceptée la transition reliant les places ViralQ4 et Ad.virales saisies qui est une transition « immédiate » et automatique servant simplement à la comptabilisation du nombre d'adresses virales fournies). Ces lois ont été précisément décrites et estimées dans le chapitre 5.

6.2.1 Synthèse des modèles à intégrer dans le réseau de Petri :

Le tableau 6.1 montre que le modèle global requiert 18 paramètres à intégrer dans le réseau de Petri. Cette vision du jeu en un système global représentée par les lois ci-dessous présente plusieurs avantages. Tout d'abord, le réseau de Petri nous permet de représenter l'ensemble du jeu alors que celui-ci est complexe. Les leviers marketing d'inscriptions (affiliation, mail annonce et relance, autopromotion du site) sont en particulier intégrés, ce qui nous permettra d'effectuer plusieurs simulations de stratégies marketing différentes. D'autre part, les lois utilisées sont des lois connues appliquées fréquemment dans le domaine du marketing pour décrire le comportement du consommateur. L'objectif étant de plus de mesurer l'impact viral du jeu, on constate que le modèle global intègre bien les différentes étapes qui constituent le mécanisme viral du jeu, notamment le taux de transformation des adresses virales en joueurs. Enfin, les lois stochastiques du modèle intègrent la variable temps (pour la séquence des inscriptions et la durée entre adresse virale fournie/adresse virale joueur) qui nous permettra de simuler différentes durées de jeu et leur impact sur la performance globale du système. On peut donc conclure déjà, de par la construction du modèle, que ce type de méthode est particulièrement adapté à la modélisation d'opérations de marketing viral on line comme le jeu « le ciel est à tout le monde ».

Tableau 6.1: synthèse des modèles des transitions du réseau de Pétri

No	Transition	Modèle	Paramètres	
1	Inscription via affiliation	Weibull	α_1 β_1	$0{,}2246\ e^{-0{,}00000006\ t}$ 0,6000
2	Inscription via mail d'annonce	Weibull	α_2 β_2	$0{,}56\ e^{-0{,}000000031\ t}$ 0,3100
3	Inscription via mail de relance	Weibull	α_3 β_3	$1{,}40\ e^{-0{,}000000021\ t}$ 0,2100
4	Inscription via autopromotion sur le site	Weibull	α_4 β_4	0,0084 0,7602
5	Inscription directe source non identifiée	Weibull	α_5 β_5	0,0125 0,7225
6	Non-réponses aux questions (définit les transitions Q1, Q2, Q32, etc…)	Béta Géométrique 2	α_6 β_6 α_7 β_7	0,35 0,06 0,00128 0,0001
7	Transformation adresse virale en joueur (Transitions ViralQ4joue et ViralQ4nejouepas)	Log réciproque	α_8 β_8	0,09 0,86
8	Durée entre saisie adresse virale et inscription au jeu de l'adresse virale (Transition grise clair inscription de la figure 6.1)	Weibull	r_9 α_9	0,4831 0,1036

Les formules des différents modèles ont été détaillées dans le chapitre 5.

6.2.2 Performance du modèle :

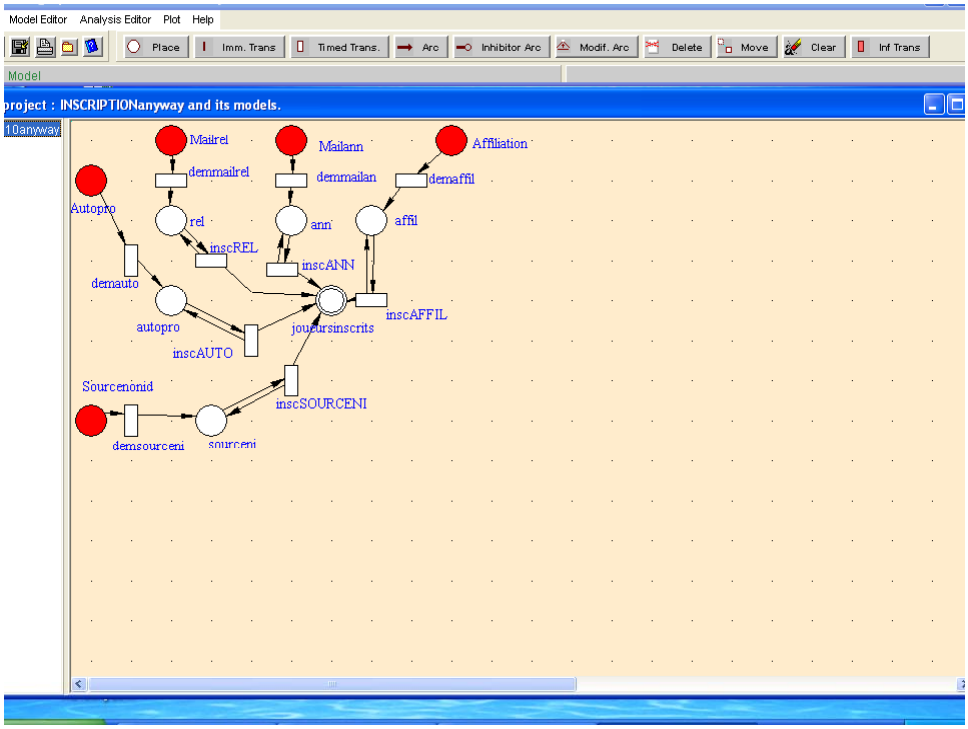
L'objectif de cette partie est d'évaluer la performance du modèle de Petri et de montrer que le modèle global représente bien les données observées. Pour des raisons de clarté de présentation et de compréhension, nous ne travaillerons dans cette partie que sur les 10 premières questions du questionnaire. En effet, nous avons dans le chapitre précédent utilisé l'ensemble des données des 41 questions pour estimer les paramètres des différentes lois. Ceci était indispensable pour nous permettre de trouver et de sélectionner des modèles robustes reposant sur de nombreuses données observées. L'objectif du chapitre 6 est différent : nous voulons modéliser le jeu Anyway pour mieux en comprendre son mécanisme et cherchons à optimiser le design de l'opération

(notamment le nombre et l'ordre des questions virales et de qualification posées) Dans une première étape (partie 6.2.2), nous allons comparer les résultats du modèle de Petri avec les données observées. Dans une deuxième étape (partie 6.3), nous effectuerons plusieurs simulations dont nous comparerons les résultats à ceux de l'opération qui a été réellement menée par Millemercis afin d'améliorer le design du jeu. Il est donc nécessaire de prendre les mêmes bases de travail pour la mesure de la performance du modèle et pour les simulations futures. Ceci explique le fait que nous allons donc travailler sur les 10 premières questions du jeu (plus de questions aurait rendu la simulation trop complexe, moins de questions limiterait les possibilités de simulation).

6.2.2.1 Performance du modèle sur la partie inscription des joueurs:

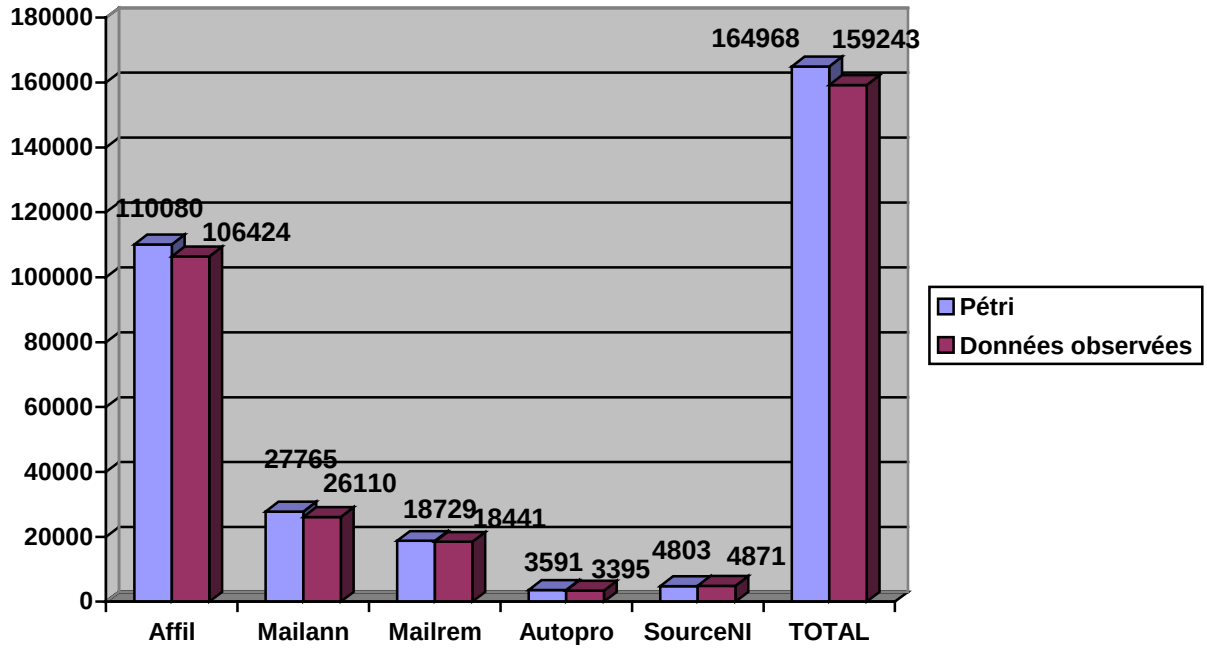
Dans une première étape, nous avons mesuré la performance du modèle de Petri sur la partie inscription des joueurs. Pour cela, nous avons fait tourner le modèle sur le système représentant uniquement les inscriptions, ce système étant présenté dans la figure 6.2 ci-dessous.

Figure 6.2 : Réseau de Petri Inscriptions



Le détail de ce réseau correspond à la description que nous en avons faite dans le chapitre 5. Rajoutons quelques précisions sur le design du réseau qui apparaît sur la figure 6.2. Tout d'abord, nous avons au départ mis un jeton dans chacune des places Affiliation, Mailann, Mailrel, Autopro, Sourcenonid. Nous allons maintenant décrire le fonctionnement du réseau sur le cas de l'affiliation (mais le principe est le même pour toutes les sources de recrutement). Le jeton de la place affiliation passe dans la place affil après 19,25h (69300 secondes) (durée de la transition demaffil) car la première inscription liée à l'affiliation commence à ce moment-là , comme nous l'avons décrit dans le chapitre 5. Depuis la place affil, le jeton passe ensuite dans la transition inscrAFFIL (selon la loi de Weibull de paramètres 0.1445 et 0.6760) jusqu'à la place joueurs inscrits (la conséquence étant qu'un nouveau joueur soit inscrit). La transition inscAFFIL renvoie également un jeton dans la place affil de manière à ce que le processus soit toujours en fonctionnement et que régulièrement, en fonction de la loi Weibull paramétrée, un jeton passe par la transition inscAFFIL puis dans la place joueursinscrits. A noter que la place joueursinscrits est représentée par un double rond, ce qui correspond à une place « fluide » permettant dans les réseaux de Petri de contenir un nombre très important (même illimité) de jetons.
Lorsqu'on fait tourner le réseau de Petri sur la partie inscription et sur la durée totale du jeu (soit 30 jours), on obtient les résultats suivants.

Figure 6.3: Performance du réseau de Pétri
Inscriptions



Les résultats sous forme de tableaux sont présentés ci-dessous :

Tableau 6.2: Performance du réseau de Pétri (inscriptions sur 30 jours)

	Nbre d'inscriptions Pétri	Nbre d'inscriptions Données observées	Pétri/Données observées
Affiliation	110080	106426	1,034
Mail annonce	27765	26110	1,063
Mail relance	18729	18441	1,016
Autopromotion	3591	3395	1,058
Source non identifiée	4803	4871	0,986
TOTAL	164968	159243	1,036

On constate une bonne performance du réseau de Petri. En effet, l'écart total est de 3,6% entre les inscriptions estimées par le réseau de Petri et les données observées sur l' ensemble du jeu pendant 30 jours (inscriptions hors marketing viral qui est une variable endogène au système). De plus, chaque source importante de recrutement est bien estimée : on a un écart maximum de 6,34% sur les principales sources des recrutement (affiliation, mailannonce, mail de relance, autopromotion et source non identifiée).

Dans un deuxième temps, nous avons mesuré la performance du réseau de Petri pour représenter les inscriptions en cours de jeu. Nous avons donc fait tourner le modèle pour 10 jours de jeu et 20 jours de jeu.

Tableau 6.4: Performance du réseau de Pétri pour 10 jours de jeu

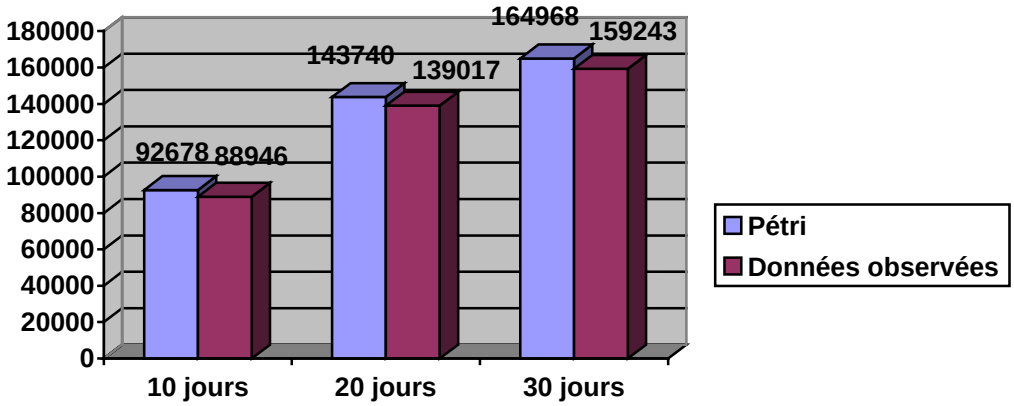
inscriptions 10 jours	Pétri	Observé	Pétri/observé
affiliation	50655	48992	1,034
mail annonce	27182	25182	1,079
mail relance	0	0	1,000
autopromo	839	821	1,022
source non identifée	2555	2404	1,063
TOTAL	81231	77399	1,050

Tableau 6.5: Performance du réseau de Pétri pour 20 jours de jeu

inscriptions 20 jours	Pétri	Observé	Pétri/observé
affiliation	91792	86892	1,056
mail annonce	27765	25983	1,069
mail relance	11769	11097	1,061
autopromo	2222	2271	0,978
source non identifée	3813	3720	1,025
TOTAL	137361	129963	1,057

On constate que le modèle représente correctement les données observées pour 10 jours et 20 jours, avec des écarts totaux d'inscriptions de l'ordre de 4% dans chaque cas. L'écart provient d'une surestimation des inscriptions venant du mail d'annonce (+6% pour 10 jours, +4% sur 20 jours), de l'affiliation (3%) qui sont les principales sources de recrutement.
La figure 6.4 donne une synthèse de ces résultats.

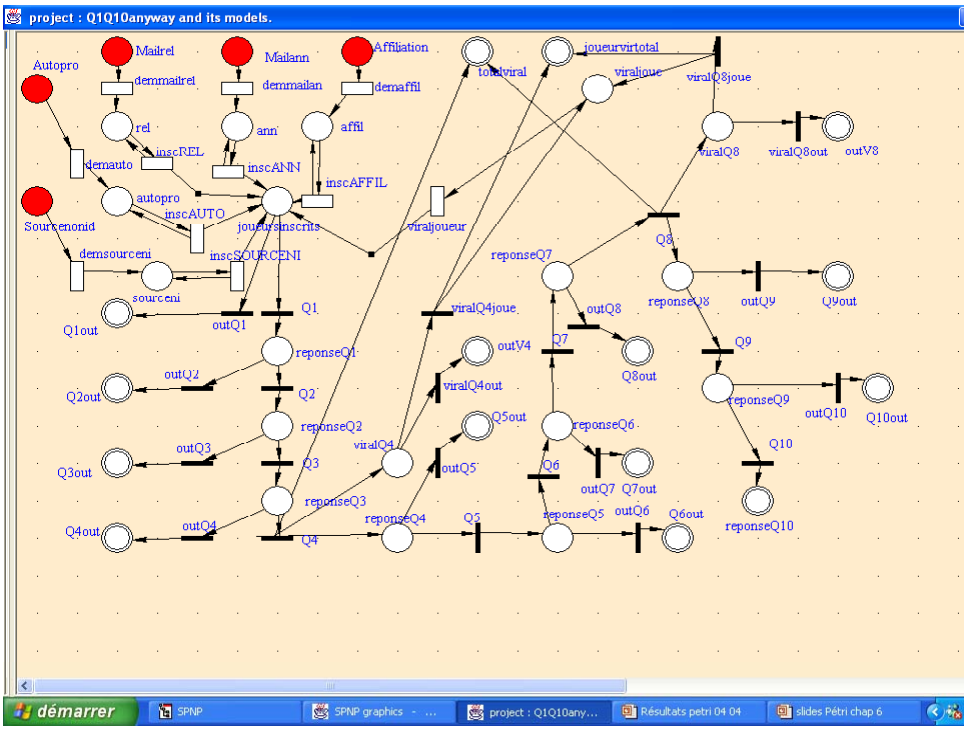
Figure 6.4: Comparaison Pétri / Données observées
à 10 jours, 20 jours, 30 jours de jeu



6.2.2.2 Performance globale du modèle

Dans une deuxième temps, nous cherchons à valider le modèle de Petri sur l'ensemble du système. Pour cela, nous avons créé le réseau tel que nous l'avons défini précédemment et incluant les modèles présentés au chapitre 5. La figure 6.4 ci-dessous représente le jeu créé sur le logiciel SPNP des réseaux de Petri.

Figure 6.5 : Opération « Le Ciel est à tout le monde » modélisé par un réseau de Petri



Chacune des transitions du modèle intègre les modèles présentés dans le tableau 6.1 et détaillés dans le chapitre 5. La partie en haut à gauche décrit les inscriptions au jeu, détaillées pour chaque source de recrutement. Les joueurs inscrits suivent alors le processus détaillé dans la partie 6.2, à savoir qu'ils sont soumis séquentiellement aux questions. Pour chaque question, soit ils répondent et passent à la question suivante, soit ils s'arrêtent (ceci correspond aux places Q1out, Q2out, etc…). Pour les questions virales, on intègre bien dans le modèle les adresses virales qui deviennent joueurs. On représente comme variable endogène les inscriptions venant de l'effet viral du jeu, puisque les adresses virales devenus joueurs passent dans la place *joueursinscrits* et recommencent le processus du système. L'un des avantages des réseaux de Petri est de particulièrement bien représenter ce phénomène viral par cet effet « réseau » (adresses fournies d'un côté par le jeu et qui rentrent elles-même dans le jeu, ce processus se reproduisant ainsi quasiment indéfiniment).

En faisant tourner le modèle de Petri sur l'ensemble du jeu, nous avons mesuré le nombre d'adresses virales fournies et le nombre d'adresses virales ayant joué. Nous avons ensuite comparé ces résultats aux données observées, sur 3 périodes : 10 jours, 20 jours, 30 jours de jeu.

Les résultats sont fournis dans le tableau 6.6.

Tableau 6.6: Performance du réseau de Petri par rapport aux données observées

10 jours de jeu	Pétri	Données observées	Pétri/Données observées
Nbre inscriptions hors viral	92678	88946	1,04
Nbre d'adresses virales	101268	111283	0,91
Nbre d'adresses virales ayant joué	20489	21248	0,96

20 jours de jeu	Pétri	Données observées	Pétri/Données observées
Nbre inscriptions hors viral	143740	139017	1,03
Nbre d'adresses virales	158203	175781	0,90
Nbre d'adresses virales ayant joué	32643	36483	0,89

30 jours de jeu	Pétri	Données observées	Pétri/Données observées
Nbre inscriptions hors viral	164968	159243	1,04
Nbre d'adresses virales	183471	201616	0,91
Nbre d'adresses virales ayant joué	39487	43392	0,91

Pour l'ensemble du jeu (soit 30 jours), on constate que le modèle de Petri donne globalement une bonne prévision du nombre d'adresses virales et du nombre de joueurs issus de ces adresses virales, égale respectivement à 91% des données observées. Pour 10 jours de jeu, le modèle est meilleur sur la prévision du nombre de joueurs transformés.

Les écarts du modèle de Petri par rapport aux données observées s'expliquent essentiellement par deux phénomènes :

1. Dans le réseau de Petri, les joueurs issus du viral re-rentrent dans le système de réponse aux questions du jeu (par la transition *viraljoueur* de la figure 6.5). Par

conséquent, l'effet de l'écart entre les estimations et les données réelles pour le modèle de transformation (des adresses virales en joueurs) et pour le modèle Béta Géométrique se multiplie au fur et à mesure que les adresses virales jouent et redonnent des adresses virales dont une partie joue. L'une des solutions est d'avoir des modèles (log réciproque ou béta géométrique ou éventuellement d'autres approches) les plus proches possible des données observées.

2. Dans le modèle de transformation des adresses virales en joueurs, nous avons considéré les paramètres stables dans le temps. Il est probable que ces paramètres ne soient pas les mêmes en début et en fin de jeu.

Après avoir mesuré la performance du modèle de Petri , nous allons désormais mener plusieurs simulations permettant d'étudier d'autres stratégies possibles de jeu.

6.3 SIMULATIONS

L'un des avantages principaux des réseaux de Petri est de pouvoir mener des simulations (Haas, 2002). Les domaines d'application pour la simulation sont nombreux (Law et Kelton, 2000) :

- définition des systèmes de production,
- évaluation de systèmes d'armes militaires et de contraintes logistiques,
- détermination de protocoles pour des réseaux de communication,

- détermination de besoins en hardware et software pour un système informatique,
- organisation de services (call centers, hôpitaux, systèmes postaux),
- refonte de processus de travail,
- analyse de systèmes financiers ou économiques.

La simulation est l'une des techniques les plus utilisées en recherche opérationnelle ou en sciences de techniques de management. Lane, Mansour et Harpell (1993) montrent sur une étude longitudinale entre 1973 et 1988 que la simulation a été notée comme l'une des trois plus importantes techniques de recherche opérationnelle (les deux autres étant la programmation mathématique et les statistiques). Gupta (1997) analysa 1294 articles du journal Interfaces (l'un des journaux leaders dans les applications en recherche opérationnelle) entre 1970 et 1992 et montra que la simulation était la deuxième technique, après la programmation mathématique, parmi les 13 techniques considérées.
La validité des techniques de simulation repose sur les modèles sous-jacents utilisés. Nous avons consacré le chapitre 5 à l'élaboration de modèles robustes permettant de représenter le plus finement possible les différents éléments du système étudié. Nous avons été particulièrement vigilante à trouver le bon modèle qui nous permette ensuite d'effectuer plusieurs simulations possibles.
Dans le cadre du jeu Anyway, nous avons cherché à tester plusieurs autres types de design du jeu au travers des simulations. Ceci permet notamment de mesurer l'impact d'autres stratégies marketing qui auraient pu être utilisées, et de comparer la performance de ces autres stratégies avec celle du jeu réel qui a été effectué. Les leviers marketing sur lesquels il semble intéressant de travailler dans le cadre de ce jeu peuvent être regroupés en trois grandes catégories :

1. Leviers sur les inscriptions au jeu :

- L'affiliation : que se passerait-il si Millemercis développait des partenariats avec d'autres sites affiliés ? Quel serait l'impact sur le nombre d'adresses virales fournies et celles transformées en joueurs ?
- Le mail de relance : serait-il intéressant de faire plusieurs mails de relance plutôt qu'un seul ?

- L'autopromotion du site : si on double le recrutement par autopromotion sur le site anyway (en améliorant par exemple la bannière du jeu et en retravaillant sur les emplacements sur le site), quel serait l'impact sur la performance du jeu ?
- Que se passerait-il si on faisait moins d'affiliation et plus de mails de relance ?

2. Leviers sur le design du questionnaire :

- Comment mieux choisir le type et l'ordre des questions virales et de qualification ? Est-ce que le jeu aurait été plus performant si on avait choisi un questionnaire avec 1 question de qualification, 1 question virale?
- Si on commence par 1 question virale suivie de 3 questions de qualification plutôt que l'inverse, que se passe-t-il ?

3. Leviers sur la durée du jeu :

- Quel serait la performance d'un jeu sur 10 jours plutôt que 20 jours ou 30 jours ?
- Ne serait-il pas préférable de faire un jeu plus court ?

L'ensemble de ces questions correspond aux préoccupations quotidiennes des entreprises qui souhaitent utiliser les techniques de marketing viral sur Internet, en particulier celles qui ont décidé de mener des opérations de jeu mais qui ne savent pas comment optimiser leur action (et donc aussi leur budget).
Le champ des simulations possibles avec le modèle de Petri que nous avons développé est immense. Nous avons décidé d'effectuer un certain nombre de simulations, le choix de celles-ci s'effectuant sur la base de stratégies marketing qui nous semblent les plus pertinentes. Ces simulations portent sur les trois leviers que nous avons détaillés ci-dessus.

6.3.1 Simulations sur les inscriptions sur 3 périodes de jeu (10, 20 et 30 jours):

Nous avons mené plusieurs simulations concernant les stratégies à mener afin de stimuler le nombre d'inscriptions au jeu :

- affiliation : on double le nombre d'inscrits venant de l'affiliation en développant de nouveaux partenariats avec des sites affiliés.
- mail de relance : on effectue 2 puis 3 mails de relance au lieu d'un seul.
- autopromotion sur le site anyway : on multiplie par deux le nombre d'inscrits via autopromotion sur le site grâce à une bannière plus efficace.
- mix des techniques de recrutement : on fait deux mails de relance et l'affiliation démarre au 15ème jour du jeu (car les coûts sont élevés en raison de la rétribution du trafic aux sites).

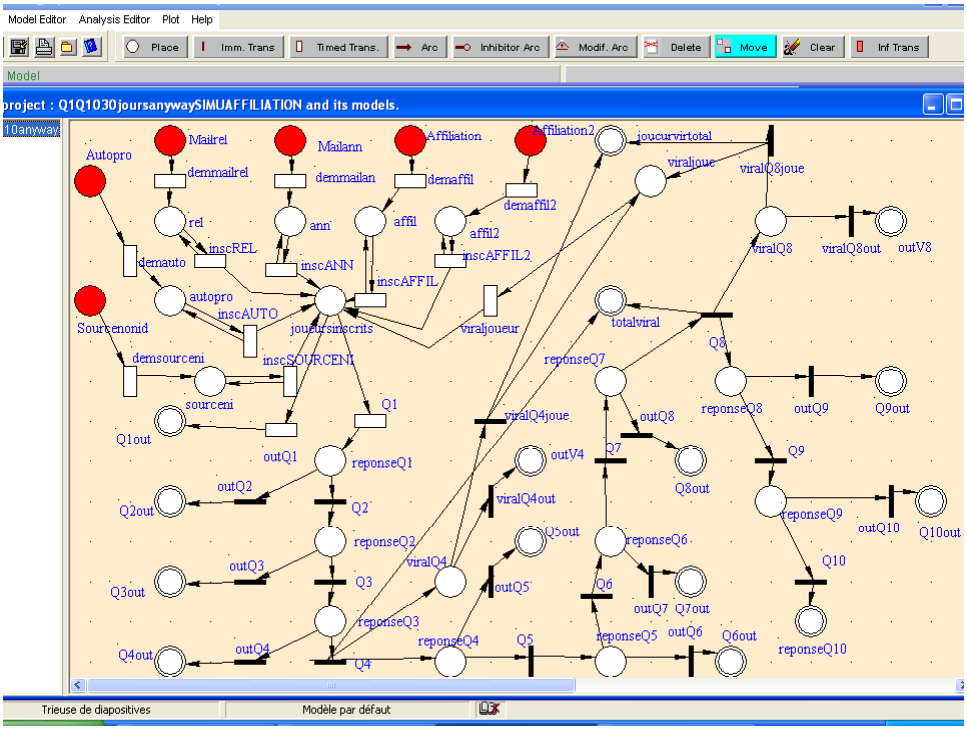
Nous avons effectué ces simulations sur trois périodes de jeu différentes, afin de croiser l'effet du temps et l'effet des stratégies différentes.

6.3.1.1 Simulation sur l'affiliation :

Pour la première stratégie (affiliation multipliée par deux) nous avons doublé dans le réseau de Petri la partie inscription par affiliation.

Ceci est représenté dans le figure 6.6. La place *Affiliation2* correspond donc au nouveau recrutement lié à l' affiliation. La transition *demaffil2* est paramétrée de la même manière que la première affiliation (transition de durée 69500 secondes pour représenter le fait que l'affiliation ne démarre qu'après cette durée). Les hypothèses n'ont pas été modifiées pour cette simulation, mais on pourrait envisager d'autres cas dans d'autres simulations. La transition *inscAFFIL2* est paramétrée avec une loi de Weibull de paramètres précisés dans le tableau 6.3 et correspondant à la période considérée (soit 10, 20 ou 30 jours).

Figure 6.6 : Réseau de Petri pour la simulation doublement du recrutement par affiliation



Le doublement des recrutements sur l'affiliation donne les résultats suivants pour les trois durées de jeu considérées (10, 20 et 30 jours):

Tableau 6.7 : Résultats de la simulation avec doublement des inscriptions affiliation

10 jours de jeu	Affiliation doublée	Jeu réel estimé par Pétri	Affiliation doublée/ Jeu réel
Nbre inscriptions hors viral	154034	92678	1,66
Nbre d'adresses virales	117990	95160	1,24
Nbre d'adresses virales ayant joué	25950	20489	1,27

20 jours de jeu	Affiliation doublée	Jeu réel estimé par Pétri	Affiliation doublée/ Jeu réel
Nbre inscriptions hors viral	238196	143740	1,66
Nbre d'adresses virales	221740	148280	1,50
Nbre d'adresses virales ayant joué	48760	32643	1,49

30 jours de jeu	Affiliation doublée	Jeu réel estimé par Pétri	Affiliation doublée/ Jeu réel
Nbre inscriptions hors viral	275048	164968	1,67
Nbre d'adresses virales	281870	170680	1,65
Nbre d'adresses virales ayant joué	62103	37352	1,66

Le tableau 6.9 montre les résultats de cette stratégie comparée à celle du jeu réel effectué par Millemercis et Anyway. On constate que doubler le recrutement par affiliation procure une hausse sur 30 jours de jeu de 65% du nombre d'adresses virales fournies et de 66% du nombre d'adresses virales transformées en joueurs. On peut conclure également qu'il n'est peut être pas nécessaire de laisser le jeu pendant 30 jours puisqu'on obtient déjà en 20 jours une hausse de 50% du nombre d'adresses virales fournies et de joueurs « viraux » (venant d'adresses virales).

6.3.1.2 Simulation sur le mail de relance

Nous avons testé ensuite l'impact sur les performances du jeu de plusieurs mails de relance envoyés, plutôt qu'un seul mail de relance comme cela a été le cas dans le jeu mené par Millemercis et anyway.

Nous avons donc simulé tout d'abord l'envoi de 2 mails de relance (au 9ème jour et au 18ème jour) puis l'envoi de 3 mails de relance (9ème jour, 18ème jour et 24ème jour). Nous avons également sur ces stratégies comparé des durées de jeu différentes : 10 jours, 20 jours ou 30 jours de jeu.

Pour simuler le 2ème mail de relance, nous avons modifié le réseau de Petri initial en ajoutant une place et des transitions correspondant au 2ème mail de relance, utilisant le même principe que le mail de relance déjà programmé initialement dans le réseau et qui représentait le mail envoyé au 18ème jour du jeu. Le processus sera le même pour le 3ème mail de relance prévu le 24ème jour du jeu. Les réseaux de Petri correspondant sont représentés dans les figures 6.7 et 6.8.

Figure 6.7 : Réseau de Petri correspondant à 2 mails de relance

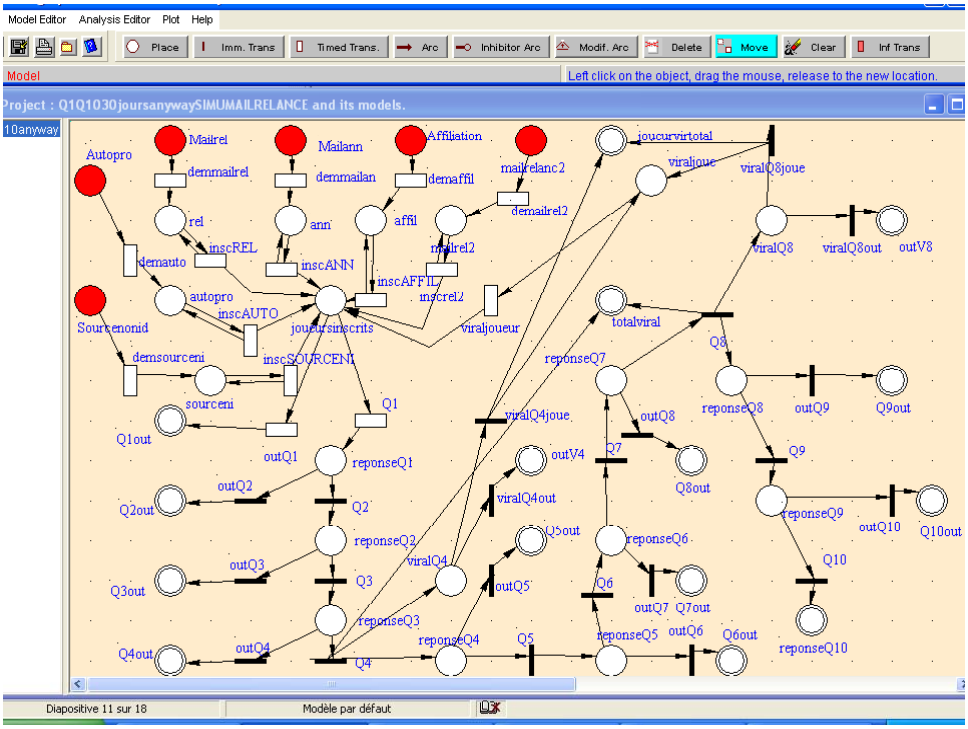
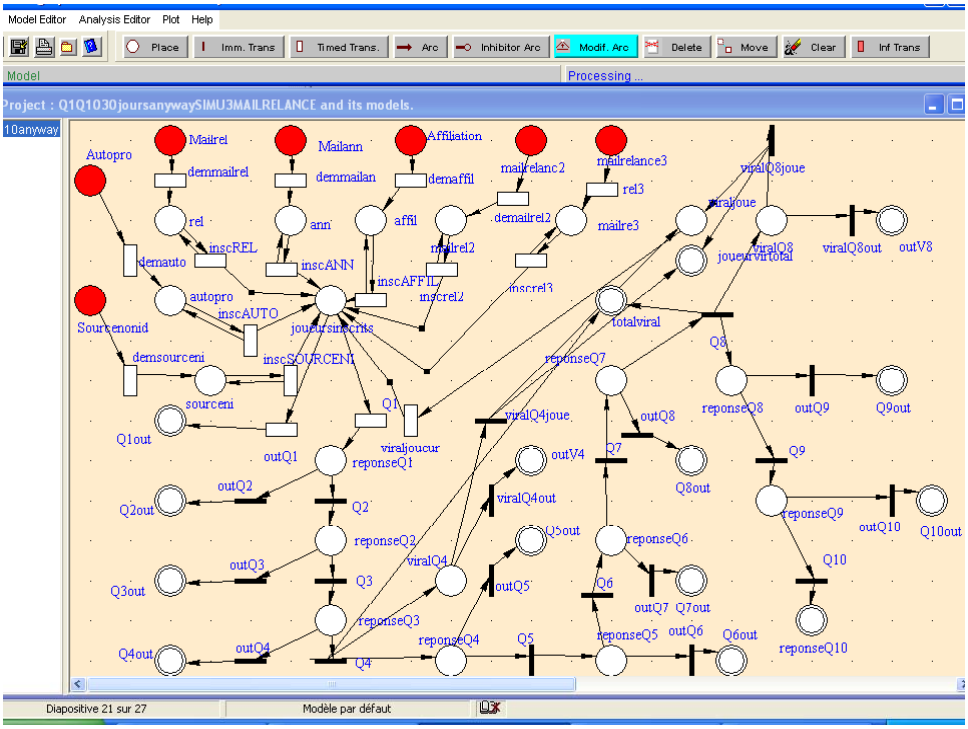


Figure 6.8 : Réseau de Petri correspondant à 3 mails de relance



Les résultats de ces simulations sont fournies dans le tableau 6.8 :

Tableau 6.8 : Simulations stratégies de mails de relance

10 jours de jeu	2 mails de relance (9ème et 18ème jour)	3 mails de relance (9ème, 18ème, 24ème jours)	Jeu réel estimé par Pétri	2 mails / jeu réel	3 mails / jeu réel	3 mails / 2 mails
Nbre d'adresses virales	96140	96140	95160	1,01	1,01	1,00
Nbre d'adresses virales ayant joué	21120	21120	20489	1,03	1,03	1,00

20 jours de jeu	2 mails de relance (9ème et 18ème jour)	3 mails de relance (9ème, 18ème, 24ème jours)	Jeu réel estimé par Pétri	2 mails / jeu réel	3 mails / jeu réel	3 mails / 2 mails
Nbre d'adresses virales	218900	218510	148280	1,48	1,47	1,00
Nbre d'adresses virales ayant joué	48209	48168	32643	1,48	1,48	1,00

30 jours de jeu	2 mails de relance (9ème et 18ème jour)	3 mails de relance (9ème, 18ème, 24ème jours)	Jeu réel estimé par Pétri	2 mails / jeu réel	3 mails / jeu réel	3 mails / 2 mails
Nbre d'adresses virales	220650	225430	170680	1,29	1,32	1,02
Nbre d'adresses virales ayant joué	48700	49760	37352	1,30	1,33	1,02

On constate dans le tableau 6.8 que la meilleure des stratégies proposées est le fait d'envoyer 2 mails de relance et que cette stratégie est particulièrement performante sur 20 jours de jeu. En effet, on constate tout d'abord que l'impact des 2ème et 3ème mails est évidemment quasi nul sur 10 jours de jeu, puisque le 2ème mail est envoyé le 9ème jour et le 3ème mail le 24ème jour du jeu. Pour une durée de 20 jours de jeu, on augmente par rapport au jeu réel le nombre d'adresses virales fournies et transformées de 48% lorsqu'on fait un 2ème mail. En revanche, l'apport du 3ème mail est évidemment nul sur 20 jours puisque le 3ème mail a lieu le 24ème jour. Enfin, sur 30 jours de jeu, un 2ème mail au 9ème jour permet d'accroître la performance globale de 29% et 3 mails de relance envoyés augmentent la performance de 32%. L'apport du 3ème mail est donc de + 2% sur la performance globale du jeu.

Lorsqu'on compare les tableaux 6.7 et 6.8 sur la période de 20 jours, il est intéressant de constater que la stratégie de 2 mails de relance (stratégie à quasi zéro coût) est aussi performante que le doublement de l'affiliation qui en revanche entraîne une hausse du budget de par la commission à rétribuer aux sites affiliés. Il est donc plus intéressant pour une période plus courte de jeu (20 jours) d'effectuer 2 mails de relance plutôt que doubler l'affiliation.

6.3.1.3 Simulation sur l'autopromotion

En ce qui concerne l'autopromotion, on a aussi testé une croissance de recrutement par cette source, en supposant cette stratégie possible grâce à une amélioration de la bannière sur le site Anyway, et à une optimisation de son emplacement sur le site (qui pourrait être obtenue par technique de eye-tracking par exemple, utilisée par les entreprises pour optimiser le design de leur site Web). L'intérêt de cette stratégie est essentiellement financier puisque l'autopromotion a un coût très faible pour l'entreprise (coût de la constitution des bannières)

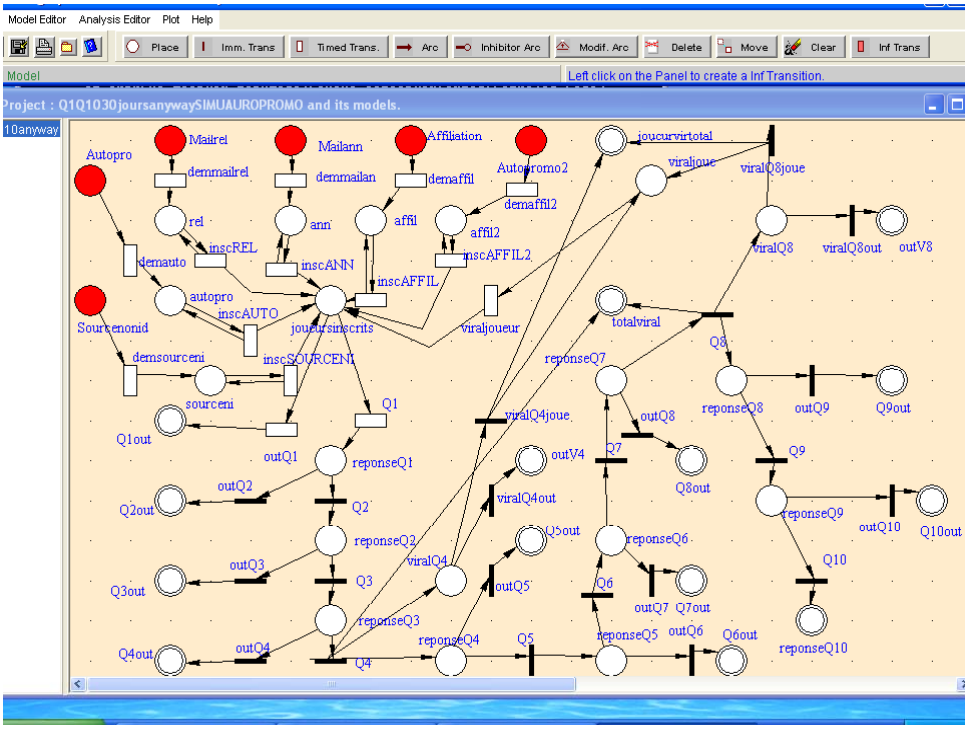
par rapport à la stratégie d'affiliation dans laquelle le site affilié touche une commission (environ 5%) sur les visites issues de ce site.

Etant donné le faible nombre d'inscriptions venant de cette source de recrutement (1%), nous avons simulé de doubler les inscriptions liées à l'autopromotion sur le site Anyway. Nous avons effet dans le jeu 3395 inscriptions provenant de l'autopromotion. Doubler ces inscriptions reviendrait donc, si on suppose que le taux de click moyen sur une bannière est de 2%, à insérer 169 750 bannières supplémentaires, ce qui est tout à fait envisageable au niveau opérationnel.

Pour effectuer cette simulation, nous avons procédé comme pour l'affiliation en créant une autre source de recrutement avec l'autopromotion, avec les mêmes paramètres qu'initialement.

Le réseau de Petri est le suivant :

Figure 6.9 : Simulation autopromotion du site



Les résultats de cette stratégie pour les 3 périodes de jeu (10, 20 et 30 jours) sont fournis dans le tableau 6.12 :

Tableau 6.9 :

Résultats des simulations recrutement par autopromotion multiplié par 2

10 jours de jeu	Autopromotion x 2	Jeu réel estimé par Pétri	Autopromotion x 100/ Jeu réel
Nbre d'adresses virales	96112	95160	1,01
Nbre d'adresses virales ayant joué	20899	20489	1,02

20 jours de jeu	Autopromotion x 2	Jeu réel estimé par Pétri	Autopromotion x 100/ Jeu réel
Nbre d'adresses virales	157280	148280	1,06
Nbre d'adresses virales ayant joué	34558	32643	1,06

30 jours de jeu	Autopromotion x 2	Jeu réel estimé par Pétri	Autopromotion x 100/ Jeu réel
Nbre d'adresses virales	174470	170680	1,02
Nbre d'adresses virales ayant joué	38619	37352	1,03

Ces résultats montre un impact faible de cette stratégie sur la performance globale du jeu, puisque le nombre d'adresses virales et transformées progressent légèrement par rapport au jeu réel pour les 2 périodes de 20 jours et 30 jours.

Le doublement de l'autopromotion est donc moins performant que le doublement de l'affiliation (ceci est tout à fait logique compte tenu du nombre beaucoup plus important d'inscriptions provenant de l'affiliation). On peut cependant réfléchir à l'amélioration du taux de click par une bannière plus ciblée sur le profil socio-démographique ou le profil de comportement de voyage décrit dans les questions de qualification auxquelles es joueurs .ont répondu.

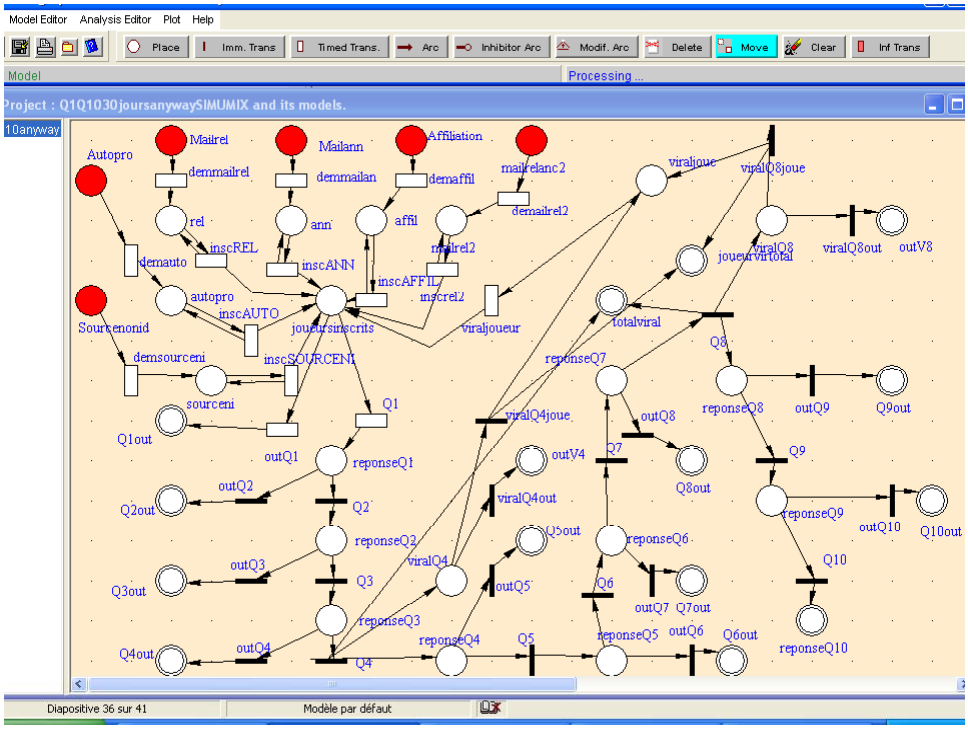
6.3.1.4 Simulation sur un mix de techniques de recrutement

Nous avons dans cette partie testé un mix de stratégies de recrutement, dans le sens où nous avons simulé une stratégie avec 2 mails de relance (9ème et 18ème jour) couplée à une affiliation diminuée de moitié car démarrant le 15ème jour du jour, dans l'objectif de réduire les coûts de l'opération liés à la rétribution des sites affiliés.

Le réseau de Petri a été modifié selon les mêmes principes que ceux présentés dans les parties 6.3.1.1 et 6.3.1.2.

Le réseau de Petri est le suivant (rajout du 2ème mail de relance et modification des paramètres de la transition *demaffil* en imposant un délai à cette transition de 15 jours, soit 1296000 secondes).

Figure 6.10 : Réseau de Petri stratégie mix de recrutement



Les résultats de cette stratégie sont les suivants :

Tableau 6.10 : Résultats de la stratégie : 2 mails de relance + affiliation démarrant le 15ème jour du jeu

10 jours de jeu	Affiliation démarrage 15ème jour 2 mails de relance	Jeu réel estimé par Pétri	mix / Jeu réel
Nbre d'adresses virales	34711	95160	0,36
Nbre d'adresses virales ayant joué	7681	20489	0,37

20 jours de jeu	Affiliation démarrage 15ème jour 2 mails de relance	Jeu réel estimé par Pétri	mix / Jeu réel
Nbre d'adresses virales	146770	148280	0,99
Nbre d'adresses virales ayant joué	32133	32643	0,98

30 jours de jeu	Affiliation démarrage 15ème jour 2 mails de relance	Jeu réel estimé par Pétri	mix / Jeu réel
Nbre d'adresses virales	147310	170680	0,86
Nbre d'adresses virales ayant joué	32407	37352	0,87

On constate ici que sur 20 jours de jeu avec 2 mails de relance et moitié moins d'affiliation (puisqu'elle commence au 15ème jour), on obtient les mêmes performances que le jeu initial.
En revanche, cette stratégie est moins performante que le jeu réel sur l'ensemble des 30 jours (indice 86). Une autre stratégie intermédiaire pourrait être de commencer l'affiliation à 8 jours par exemple, ce qui aurait augmenté les performances du jeu tout en diminuant les coûts
Cette simulation de stratégie de mix de recrutement montre que les réseaux de Petri peuvent permettre également de simuler des stratégies de baisse de coûts des entreprises, puisque la stratégie ici proposée est moins coûteuse que le jeu réellement mené.

6.3.2 Simulations sur le design du questionnaire :

Nous testons dans cette partie des stratégies de design de questionnaire différentes. Notamment, nous cherchons à mesurer l'impact de l'ordre et du type de questions posées sur la performance globale du jeu. Pour cela, nous avons choisi de garder le processus des inscriptions du jeu réellement mené par Millemercis et Anyway, afin de mesurer dans un premier temps uniquement l'impact de l'ordre et du type de question.
Le questionnaire utilisé dans le jeu réel est basé sur le principe de 3 questions de qualification suivies de 1 question virale. Nous proposons donc de tester les questionnaires suivants (toujours sur 10 questions au total):

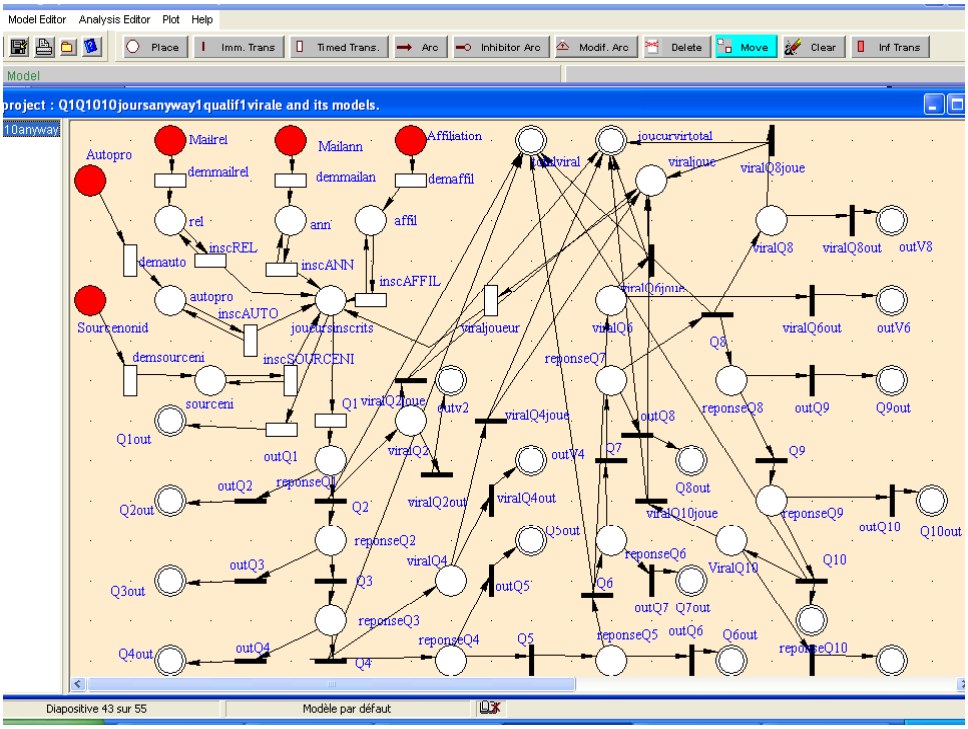
- 10 questions alternant 1 question de qualification / 1 question virale
- 10 questions alternant 2 questions de qualifications/ 2 virales
- 10 questions avec 5 questions de qualification suivies de 5 questions virales
- 10 questions avec 5 questions virales suivies de 5 questions de qualification

Les résultats de ces simulations seront comparés à la performance du jeu réel.
Nous avons mené ces simulations sur la durée complète du jeu, soit 30 jours, puis sur 10 jours et 20 jours comme pour les simulations précédentes. Seuls les paramètres des lois de Weibull concernant les inscriptions sont modifiés dans le réseau de Petri entre les différentes périodes de jeu.

6.3.2.1 Simulation 10 questions alternant 1 question de qualification / 1 question virale

Nous avons dans un premier temps testé un questionnaire alternant sur les 10 questions
1 question de qualification / 1 question virale, afin de tester son impact en terme de performance. Le réseau de Petri a donc été modifié en paramétrant les questions 1, 3, 5, 7, et 9 comme questions de qualification et les questions 2, 4, 6, 8, 10 comme questions virales. Nous avons appliqué le modèle de réponse béta géométrique 2 du chapitre 5 pour estimer les probabilités d'arrêt à chacune de ces questions. Les autres paramètres du système (notamment les inscriptions) correspondent à ceux utilisés pour modéliser le jeu réel.
Le réseau de Petri correspondant est le suivant :

Figure 6.11 : Réseau de Petri stratégie 1 question de qualification / 1 question virale



Les résultats de la simulation de ce type de questionnaire sont les suivants :

Tableau 6.11 : Résultats simulation 1 question de qualification / 1 question virale

10 jours de jeu	1 quest. qualif / 1 quest. Virale	Jeu réel estimé par Pétri	1 qualif / 1 virale vs Jeu réel
Nbre inscriptions hors viral	92678	92678	1,00
Nbre d'adresses virales	144580	95160	1,52
Nbre d'adresses virales ayant joué	33805	20489	1,65

20 jours de jeu	1 quest. qualif / 1 quest. Virale	Jeu réel estimé par Pétri	1 qualif / 1 virale vs Jeu réel
Nbre inscriptions hors viral	143740	143740	1,00
Nbre d'adresses virales	223130	148280	1,50
Nbre d'adresses virales ayant joué	52193	32643	1,60

30 jours de jeu	1 quest. qualif / 1 quest. Virale	Jeu réel estimé par Pétri	1 qualif / 1 virale vs Jeu réel
Nbre inscriptions hors viral	164968	164968	1,00
Nbre d'adresses virales	256470	170680	1,50
Nbre d'adresses virales ayant joué	59760	37352	1,60

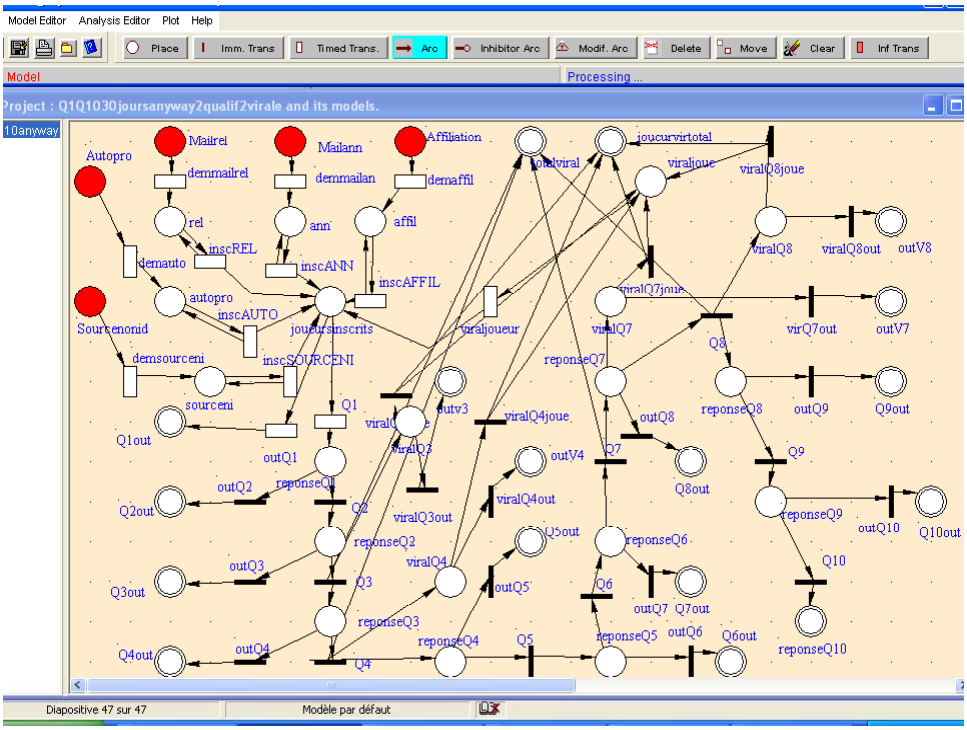
De par le fait que le questionnaire simulé comporte 5 questions virales au lieu de 2 questions virales (questions 4 et 8) dans le jeu réel, on s'attend à ce que le nombre d'adresses virales fournies et transformées soient plus important que dans le jeu réel. C'est effectivement le cas, comme on peut le constater sur le tableau 6.14. Cependant, on ne double pas le nombre d'adresses virales fournies, même si le nombre de questions virales est plus du double de celui du jeu réel. Que le jeu dure 10, 20 ou 30 jours, la performance d'un questionnaire alternant 1 question de qualification / 1 question virale est 50% plus forte qu'un questionnaire de 10 questions avec uniquement 2 questions virales (la 4 et la 8). Ceci s'explique par les fortes probabilités d'arrêt sur chaque question virale qui diminuent rapidement le nombre de joueurs qui répondent à plusieurs questions.

6.3.2.2 Simulation 10 questions alternant 2 questions de qualifications/ 2 virales

Nous avons approfondi les recherches sur la bonne adéquation de l'ordre et du type de questions posées. Nous avons donc dans cette partie testé un autre type de questionnaire comportant 2 questions de qualifications suivies de 2 questions virales. Sur 10 questions, les questions 1, 2, 5, 6, 9 et 10 sont donc des questions de qualification et les questions 3,4, 7 et 8 qui sont des questions virales. Le nombre de questions virales est donc moins important que dans la simulation précédente (4 au lieu de 5). En revanche, 2 questions virales se suivent à chaque fois.

Le réseau de Petri est le suivant :

Figure 6.12 : Réseau de Petri 2 questions de qualification / 2 questions virales



Les résultats de ces simulations sont les suivants :

Tableau 6.12 : Résultats simulation 2 questions de qualification / 2 questions virales

10 jours de jeu	2 quest. qualif / 2 quest. Virale	Jeu réel estimé par Pétri	2 qualif / 2 virale vs Jeu réel
Nbre inscriptions hors viral	92678	92678	1,00
Nbre d'adresses virales	132710	95160	1,39
Nbre d'adresses virales ayant joué	32418	20489	1,58

20 jours de jeu	2 quest. qualif / 2 quest. Virale	Jeu réel estimé par Pétri	2 qualif / 2 virale vs Jeu réel
Nbre inscriptions hors viral	143740	143740	1,00
Nbre d'adresses virales	209620	148280	1,41
Nbre d'adresses virales ayant joué	44006	32643	1,35

30 jours de jeu	2 quest. qualif / 2 quest. Virale	Jeu réel estimé par Pétri	2 qualif / 2 virale vs Jeu réel
Nbre inscriptions hors viral	164968	164968	1,00
Nbre d'adresses virales	242860	170680	1,42
Nbre d'adresses virales ayant joué	50997	37352	1,37

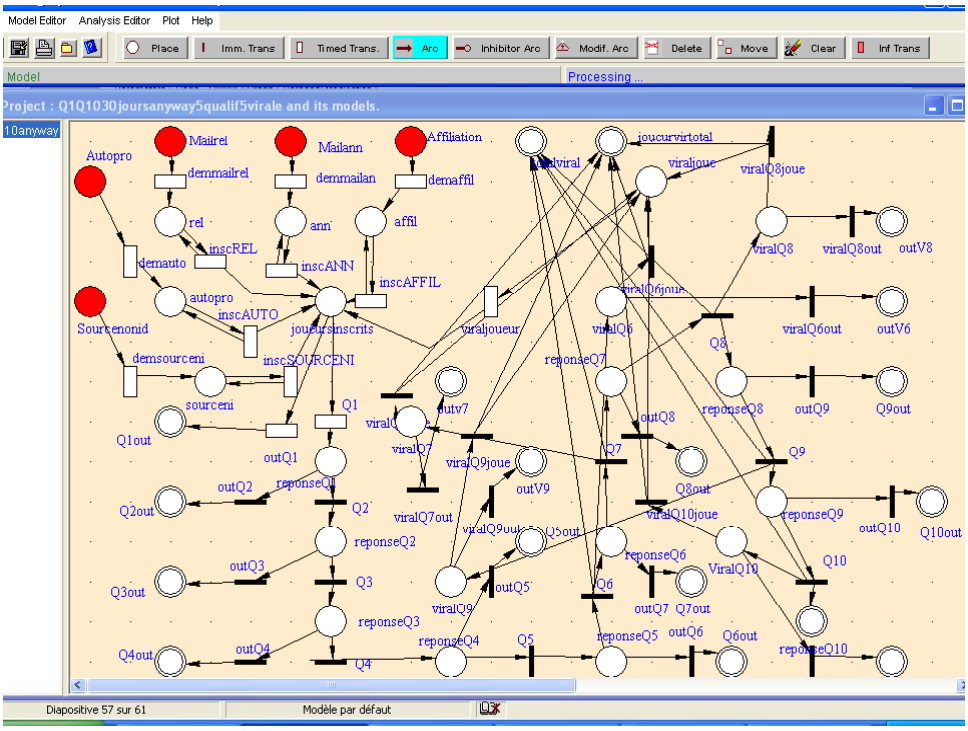
On constate dans ces résultats que quelle que soit la durée du jeu, la performance d'un questionnaire 2 questions de qualification / 2 questions virales est de 40 % plus forte (pour le nombre d'adresses virales fournies) que le jeu qui a été mené dans la réalité. En ce qui concerne les adresses virales transformées en joueurs, la performance est accrue de 37% (sur 30 jours). On peut également noter la bonne performance de ce questionnaire par rapport à la stratégie précédente de 1 question de qualification / 1 question virale (comparaison des tableaux 6.11 et 6.12).

6.3.2.3 Simulation 10 questions avec 5 questions de qualification suivies de 5 questions virales

Nous avons cherché également à tester ici un questionnaire sans alternance questions de qualification / questions virales, mais plutôt un questionnaire dans lequel toutes les questions de qualification sont posées au début et toutes les questions virales à la fin.
Nous avons donc simulé questions de qualification suivies de 5 questions virales.

Le réseau de Petri correspondant est le suivant :

Figure 6.13 : Réseau de Petri 5 questions de qualification / 5 questions virales



Les résultats sont les suivants :

Tableau 6.13 : Résultats des simulations questions de qualification / 5 questions virales

10 jours de jeu	5 quest. qualif / 5 quest. Virale	Jeu réel estimé par Pétri	5 qualif / 5 virale vs Jeu réel
Nbre inscriptions hors viral	92678	92678	1,00
Nbre d'adresses virales	119480	95160	1,26
Nbre d'adresses virales ayant joué	24377	20489	1,19

20 jours de jeu	5 quest. qualif / 5 quest. Virale	Jeu réel estimé par Pétri	5 qualif / 5 virale vs Jeu réel
Nbre inscriptions hors viral	143740	143740	1,00
Nbre d'adresses virales	184880	148280	1,25
Nbre d'adresses virales ayant joué	37905	32643	1,16

30 jours de jeu	5 quest. qualif / 5 quest. Virale	Jeu réel estimé par Pétri	5 qualif / 5 virale vs Jeu réel
Nbre inscriptions hors viral	164968	164968	1,00
Nbre d'adresses virales	214650	170680	1,26
Nbre d'adresses virales ayant joué	44023	37352	1,18

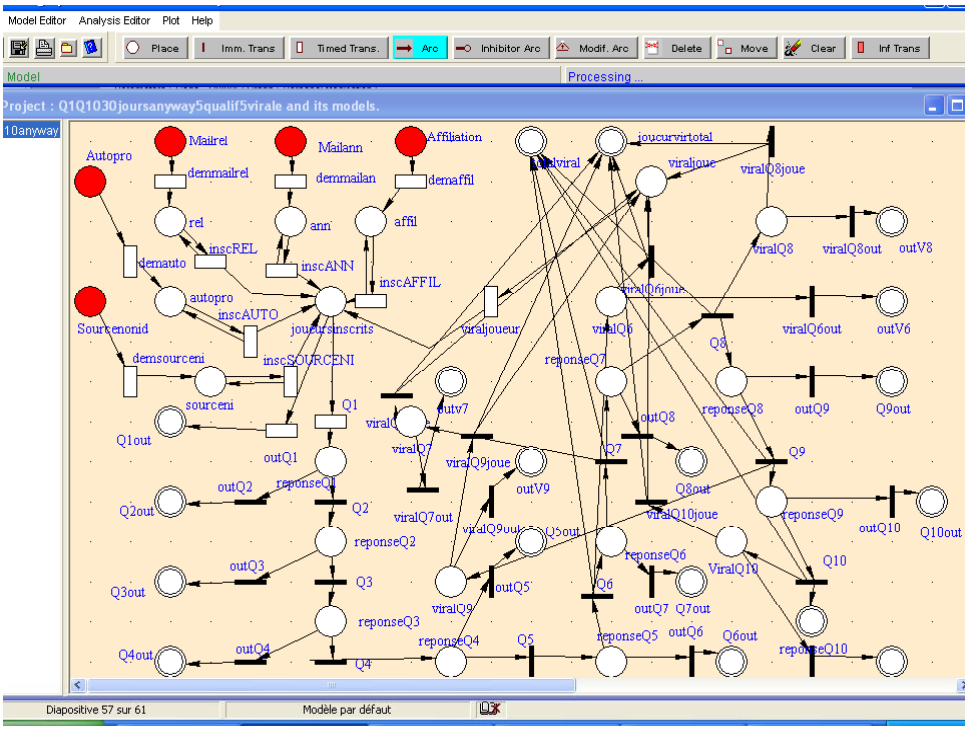
Ces résultats montrent une performance accrue de 26% par rapport au jeu réel sur le nombre d'adresses virales fournies et de 18% sur le nombre d'adresses virales transformées en joueurs lorsqu'on utilise un questionnaire avec 5 questions de qualification suivies de 5 questions virales. Cette stratégie est intéressante car elle est fondée sur un nombre de renseignements minimum que l'entreprise souhaite récolter sur l'internaute, suivie ensuite de la constitution d'un fichier d'adresses virales de prospection.

6.3.2.4 Simulation 10 questions avec 5 questions virales suivies de 5 questions de qualification

L'ordre des questions de qualification et des questions virales est une problématique intéressante. C'est la raison pour laquelle nous avons aussi simulé la même stratégie que précédemment (soit 5 questions virales et 5 questions de qualification) mais en échangeant l'ordre: les 5 questions virales sont en premier, suivies de 5 questions de qualification. Nous supposons ici que les paramètres du modèle restent les mêmes lorsqu 'on adopte cette stratégie.

Le réseau de Petri est le suivant :

Figure 6.14 : Réseau de Petri 5 questions virales / 5 questions de qualification



Les résultats sont fournis dans le tableau ci-dessous :

Tableau 6.14 : Résultats simulation 5 questions virales/ 5 questions de qualification

10 jours de jeu	5 quest.virales / 5 quest. qualif.	Jeu réel estimé par Pétri	5 virales / 5 qualif. vs Jeu réel
Nbre inscriptions hors viral	92678	92678	1,00
Nbre d'adresses virales	274670	95160	2,89
Nbre d'adresses virales ayant joué	55568	20489	2,71

20 jours de jeu	5 quest.virales / 5 quest. qualif.	Jeu réel estimé par Pétri	5 virales / 5 qualif. vs Jeu réel
Nbre inscriptions hors viral	143740	143740	1,00
Nbre d'adresses virales	425060	148280	2,87
Nbre d'adresses virales ayant joué	85989	32643	2,63

30 jours de jeu	5 quest.virales / 5 quest. qualif.	Jeu réel estimé par Pétri	5 virales / 5 qualif. vs Jeu réel
Nbre inscriptions hors viral	164968	164968	1,00
Nbre d'adresses virales	485930	170680	2,85
Nbre d'adresses virales ayant joué	98005	37352	2,62

Les résultats de cette simulation montrent un gain important de performance puisque ce design de questionnaire permet d'améliorer la performance globale de 85% sur le nombre d'adresses virales fournies et de 62% sur le nombre d'adresses virales transformées, ces résultats étant similaires pour les trois durées de jeu.

Cette stratégie montre également l'une des limites des techniques de simulation. Nous avons supposé que les paramètres étaient les mêmes que dans le jeu réel, mais on peut se demander toutefois si dans ce cas extrêmes ils sont différents. Etant donné que les résultats de cette stratégie sont particulièrement intéressants, on pourrait mener une expérience en appliquant cette stratégie sur le même jeu afin de mesurer si justement les paramètres du modèle bougent ou non.

6.3.2.5 Synthèse des simulations testées et implications managériales

Il est intéressant de comparer toutes les simulations menées afin de déterminer les stratégies les plus performantes. Les réseaux de Petri sont utilisés ici comme système d'aide à la décision et répondent à la question : quelle est la meilleure stratégie parmi les options testées ?

Le tableau 6.15 présente une synthèse des simulations qui sont comparées en terme de performance à celle du jeu réel qui sert de base 100.

Tableau 6.15: Synthèse des simulations

	Nbre adresses virales Simulation	Nbre adresses virales jeu réel	Nbre d'adresses virales ayant joué Simulation	Nbre d'adresses virales ayant joué Jeu réel
Affiliation doublée				
10j	124	100	127	100
20j	150	100	149	100
30j	165	100	166	100
2 mails relance				
10j	101	100	103	100
20j	148	100	148	100
30j	129	100	130	100
3 mails relance				
10j	101	100	103	100
20j	147	100	148	100
30j	132	100	133	100
autopromotion x 2				
10j	101	100	102	100
20j	106	100	106	100
30j	102	100	103	100
2 mails relance / affiliation 15ème jour				
10j	36	100	37	100
20j	99	100	98	100
30j	86	100	87	100
1 qualif / 1 virale				
10j	152	100	165	100
20j	150	100	160	100
30j	150	100	160	100
2 qualif/ 2 virales				
10j	139	100	158	100
20j	141	100	135	100
30j	142	100	137	100
5 qualif/ 5 virales				
10j	126	100	119	100
20j	125	100	116	100
30j	126	100	118	100
5 virales/5 qualif				
10j	289	100	291	291
20j	287	100	263	263

Plusieurs constats intéressants et implications managériales peuvent être tirés de ce tableau :

- La meilleure stratégie sur 30 jours est de doubler l'affiliation qui permet d'accroître de 65% le nombre d'adresses virales fournies et le nombre d'adresses virales converties en joueurs,

- La 2ème stratégie performante sur 30 jours consiste à modifier l'ordre et le type de questions et de proposer une question virale suivie d'une question de qualification. On augmente dans ce cas la performance du jeu de 50%,

- La 3ème stratégie en terme de performance est l'alternance 2 questions de qualification/ 2 questions virales (+40% d'adresses virales sur 30 jours),

- Viennent ensuite les stratégies de doublement ou triplement de l'e-mail de relance. On constate que ces deux stratégies permettent d'accroître le nombre d'adresses virales de 29%. Notons également que la stratégie de 2 mails de relance est aussi performante sur 20 jours de jeu que le doublement de l'affiliation qui est coûteux. Cette stratégie est donc à développer, puisque l'envoi d'e-mails représente un coût supplémentaire très faible pour l'entreprise,

- La même conclusion peut être tirée de la stratégie de 3 mails de relance. On constate cependant que l'apport du 3ème mail par rapport au 2ème est faible (+ 3% sur 30 jours),

- Le fait de doubler le recrutement par autopromotion sur le site Anyway n'améliore pas fortement la performance globale du jeu. Cette stratégie est de plus limitée puisqu'un doublement de ce recrutement requiert un nombre de bannières supplémentaires très important (compte tenu du taux de clic moyen de 2%).De plus, l'autopromotion ne correspond pas complètement au principal objectif de l'opération qui est de constituer une base d'e-mail de prospects qui ne sont pas pour l'instant clients Anyway,

- En diminuant l'affiliation (démarrage uniquement le 15ème jour) et en envoyant 2 mails de relance, on diminue la performance par rapport au jeu réel sur 30 jours (indice 86). En revanche, cette stratégie est particulièrement intéressante sur 20 jours puisqu'on

obtient les mêmes résultats que le jeu réel avec des coûts moindres (la rétribution des sites affiliés dans cette stratégie est diminuée par 2). Il serait éventuellement intéressant de creuser plus précisément cette stratégie en effectuant d'autres simulations (affiliation qui démarre après la première semaine de jeu par exemple),

- En terme d'ordre et de choix de type de questions, on constate que la meilleure stratégie (parmi celles testées) est un questionnaire avec 5 questions virales/ 5 questions de qualification, qui permet de multiplier par 2.8 le nombre d'adresses virales fournies. Cette stratégie est préférable à l'ordre inverse où on commence le questionnaire avec 5 questions de qualification suivies de 5 questions virales. Cependant, cette stratégie est à relativiser car on peut se demander si les paramètres du modèle proposé sont encore valides pour ce cas « extrême ». Il faudrait donc pour ce cas faire plus de tests, construire une base données, afin de vérifier et mesurer si les paramètres sont différents.

CONCLUSION :

L'objectif de ce chapitre était de mesurer la performance des réseaux de Petri puis d'effectuer des simulations. Nous avons dans un premier temps synthétisé l'ensemble des modèles représentant les différentes parties du jeu. Nous avons en effet trois grands modèles intégrés. Le premier modèle concerne les inscriptions au jeu. Il représente les flux d'inscriptions par source (affiliation, mail d'annonce, mail de relance, autopromotion sur le site Anyway, source non identifiée) et se fonde sur des lois de Weibull dont on fait varier l'un des paramètres dans le temps pour les principales sources de recrutement. Le deuxième modèle est celui qui représente les probabilités d'arrêt pour chaque question du questionnaire. Il se fonde sur un modèle béta géométrique détaillé dans le chapitre 5. Enfin, le troisième modèle est un modèle de conversion qui représente la transformation des adresses virales en joueurs. Ce modèle est l'adaptation du modèle log réciproque.

Nous avons montré que l'intégration de ces modèles dans le modèle global de Petri permet d'obtenir de bonnes performances en termes de validation. Ces performances sont particulièrement bonnes pour les inscriptions, sur le nombre d'adresses virales fournies et le nombre de joueurs issus d'adresses virales, le modèle donnant une précision de 10% par rapport aux données observées.

Dans une deuxième partie de ce chapitre, nous avons montré l'intérêt d'utiliser les réseaux de Petri pour simuler des stratégies marketing différentes. Nous avons simulé 9 stratégies (affiliation doublée, 2 mails de relance, 3 mails de relance, autopromotion multipliée par 2, 2 mails de relance et affiliation diminuée par deux, 1 qualification/ 1 virale, 2 qualifications/2 virales, 5 qualifications/ 5 virales, 5 virales/ 5 qualifications). Pour chacune de ces stratégies, nous avons simulé trois durées différentes (10 jours, 20 jours, 30 jours). Ces simulations permettent de conclure à la flexibilité des réseaux de Petri, à l'intérêt pour le marketing de ce type d'approche puisque ces simulations constituent des systèmes d'aide à la décision pour les professionnels du marketing. En effet, ils peuvent avant de lancer une opération de marketing viral sur Internet simuler plusieurs options possibles, en mesurer les performances et les coûts et sélectionner la meilleure de ces stratégies.

CONCLUSION FINALE :

L'objectif de cette thèse est d'introduire la méthodologie des réseaux de Petri pour modéliser des problématiques marketing liées aux nouvelles technologies et en particulier à Internet. Nous cherchons en effet à représenter des systèmes multi-agents interactifs sur Internet.
Dans une première partie, nous avons essayé de mieux définir le concept de système multi-agents interactifs. Nous avons pour cela mené une revue de littérature s'inspirant de courants de recherche très divers mais qui tous apportent une contribution à l'étude de ces systèmes multi-agents dans un contexte interactif. Dans un premier temps, nous avons synthétisé dans le domaine du marketing les différents articles portant sur le concept d'interactivité et avons montré que plusieurs approches cohabitent, que l'interactivité est un concept multi-dimensionnel et que de futures recherches sont encore nécessaires pour mieux comprendre et définir ce concept. Le deuxième courant que nous avons présenté concerne les approches développées en psychologie sociale qui étudient en particulier les interactions des internautes au sein des communautés virtuelles. Selon ces théories, le comportement de l'internaute est influencé par la communauté virtuelle à laquelle il appartient et des approches issues de sociologie ou même d'ethnographie sont développées pour mieux comprendre les interactions des internautes au sein de ces communautés. Nous avons souligné que ce courant de recherche est très actif en marketing et que de futures recherches restent encore à mener dans ce domaine. Nous avons également présenté une troisième approche théorique développée en stratégie, abordant l'interactivité comme source de co-production de valeur pour les entreprises et les consommateurs. La quatrième source théorique que nous avons détaillée concerne l'ensemble des recherches issues de l'intelligence artificielle et de l'informatique portant sur les systèmes multi-agents, avec une définition très précise du concept de systèmes multi-agents. Enfin, nous avons également décrit les différents modèles développés dans la littérature en marketing pour étudier et comprendre le comportement des internautes. Ce domaine est très actif et récent en terme de recherches et de publications dans notre domaine et là encore de nombreuses pistes de recherche restent ouvertes.
Cette revue de littérature nous a permis de mieux appréhender le concept de systèmes multi-agents interactifs au travers de théories diverses mais tout à fait complémentaires.
Elle montre que même si de nombreuses recherches ont été menées dans ces différents champs, l'étude des systèmes multi-agents interactifs constitue une source de nombreuses pistes de recherche futures.

Dans le deuxième chapitre de la thèse, nous avons présenté les modèles des réseaux de Pétri en détaillant tout d'abord le principe de la modélisation sur des cas simples (modèle de base d'une wish list, processus de commande d'un client). Après avoir décrit les différentes familles des réseaux de Petri, nous avons ensuite développé en particulier les réseaux de Petri stochastiques qui ont été utilisés dans cette thèse. Enfin, nous avons effectué dans ce chapitre une revue de littérature sur les réseaux de Petri synthétisant les articles appliquant cette technique sur des problématiques liées au domaine de la gestion.
Nous avons ensuite consacré les quatre chapitres suivant de la thèse (chapitres 3 à 6 inclus) à appliquer les réseaux de Petri à la modélisation de deux applications marketing. Dans les chapitres 3 et 4, nous avons modélisé un système de wish list sur Internet, les chapitres 5 et 6 étant consacrés à la modélisation d'un jeu promotionnel de marketing viral sur Internet.
L'un des objectifs de la thèse étant de relier plusieurs courants de recherche tentant de comprendre les systèmes multi-agents sur Internet et d'introduire dans le domaine du marketing une technique développée et utilisée en informatique et en recherche opérationnelle pour étudier ce thème de recherche, les deux applications que nous avons développées nous ont permis de montrer que les réseaux de Petri s'appliquent effectivement à la modélisation de problématiques marketing et que ces modèles présentent un certain nombre d'avantages que nous allons synthétiser ci-dessous.

Avantages des réseaux de Petri :

- Pour les deux applications que nous avons menées, les réseaux de Petri constituent un outil essentiel de modélisation. En effet, il était impossible de prévoir de manière intuitive les résultats des simulations, que ce soit dans le cas des systèmes de wish list (tester l'intervalle de 7 jours entre 2 e-mails versus 23 jours) ou celui du jeu viral (9 stratégies marketing ont été simulées).

- Nous avons montré que sur les deux applications que nous avons modélisées, les réseaux de Petri fournissent de bonnes estimations par rapport aux données observées sur les variables que nous avons mesurées. Ceci nous permet d'affirmer la validité de ce type de modélisation sur ces deux problématiques de marketing.

- Nous avons montré également la flexibilité du modèle en termes de simulations (simulations de 2 stratégies pour la première application et de 9 stratégies pour la 2ème application), avec possibilité notamment de combiner différentes stratégies (par exemple dans le jeu viral le démarrage de l'affiliation en milieu de jeu couplé au doublement de l'e-mail de relance) et d'effectuer également des simulations sur des périodes diverses (nous avons mené des simulations sur 10, 20 et 30 jours de jeu).

Plus globalement, les réseaux de Petri constituent un outil puissant pour construire des modèles représentant des systèmes dynamiques complexes. Ils possèdent an particulier une grande flexibilité concernant la modélisation des processus stochastiques. Nous avons en effet construit les réseaux de Petri de nos deux applications en déterminant tout d'abord les blocs de modèles décrivant les différents évènements modifiant les états du système global. Nous avons notamment utilisé des lois de Weibull, des lois Gamma, des lois NBD et des modèles plus complexes (comme par exemple le modèle béta géométrique de probabilité d'arrêt aux questions dans le jeu viral, ou encore le modèle log-réciproque de conversion des adresses virales en joueurs) que nous avons pu ensuite intégrer dans le modèle de Petri. Le support des réseaux de Petri que nous avons utilisé (le logiciel SPNP) présente aussi l'avantage d'intégrer un large éventail de distributions paramétrables (exponentiel, constante, uniforme, normale tronquée weibull, log-normale, géométrique, erlang, pareto, cauchy, beta, gamma, binomial, triangulaire, loglogistic, cox2, hyperexponentielle, hypoexponentielle, défective) et permet de paramétrer des processus complexes (pour plus de précisions, on peut se référer au manuel d'utilisation de SPNP, téléchargeable sur Internet à l'adresse http://www.ee.duke.edu/~kst/ software packages).

- Les réseaux de Petri sont aussi des modèles efficaces pour représenter des systèmes dynamiques complexes en raison du principe même de la méthode. Elle consiste à réduire la complexité du phénomène étudié et à simplifier l'analyse grâce à :
    - Une décomposition du système global en différents états et transitions qui sont chacun caractérisés par des « sous-modèles », permettant une meilleure compréhension du phénomène étudié et une modélisation précise des différents évènements,
    - Une représentation graphique simple permettant une visualisation de la globalité du système mais aussi de chacun de ses éléments. Cette

représentation graphique est un atout majeur pour l'utilisation des réseaux de Petri par les praticiens du marketing. En effet, cette représentation graphique permet une meilleure compréhension du modèle par les professionnels et un dialogue possible entre les managers de marketing et les chercheurs experts en modélisation.

- La théorie des réseaux de Petri s'appuie sur des modèles mathématiques robustes, avec des méthodes de validation fondées sur de nombreux résultats théoriques. Les méthodes de simulation dans les réseaux de Petri stochastiques se fondent en particulier sur les théories rigoureuses qui se sont développées sur les vingt-cinq dernières années pour étudier les simulations. La majeure partie de ces théories porte sur le problème de l'obtention d'estimations valides et des intervalles de confiance pour des simulations longues. Ces estimations et ces intervalles de confiance sont typiquement utilisés pour comparer des systèmes ou des processus et leur validité est donc fondamentale. Ils forment aussi la base des procédures d'optimisation basées sur de la simulation. La vérification des propriétés requises pour une simulation spécifique étant souvent une tâche compliquée, le problème de la simulation doit donc être bien posé de manière à ce que le système soit stable et que les mesures de performance sur une période longue (voire infinie) existent. Le formalisme des réseaux de Petri est suffisamment puissant pour permettre une modélisation précise d'un nombre élevé de systèmes réels et suffisamment simple pour apporter de la stabilité,

- Nous avons également montré que les réseaux de Petri sont particulièrement adaptés à la modélisation de phénomènes sur Internet. Ils sont en effet performants pour modéliser notamment:

    - Des actions parallèles et/ou synchronisées : de nombreux phénomènes sur Internet comprennent ce type d'actions. L'analyse du comportement des internautes sur un site par exemple inclut des actions menées en parallèle. Dans le jeu viral, un internaute peut répondre à la première question du questionnaire tandis qu'au même moment, un autre internaute s'inscrit ou encore une adresses virale reçoit simultanément un e-mail avant de se transformer en joueur. On retrouve des phénomènes similaires lorsqu'on cherche à étudier le comportement

général de navigation des internautes sur un site. On peut avoir par exemple en même temps des internautes arrivant sur la homepage, des internautes lisant les pages descriptives sur les produits et des internautes fournissant leur numéro de carte bancaire pour pouvoir acheter.

- Des actions séquentielles : la navigation des internautes sur un site Web est par essence séquentielle (arrivée sur la homepage, visualisation de certaines pages, achat/non achat par exemple). Nous avons décrit dans la thèse les différents recherches en marketing modélisant le comportement de navigation par des modèles séquentiels. Les réseaux de Petri permettent aussi cette modélisation d'actions séquentielles, le principe même de la méthode étant de décomposer le phénomène complexe en différents états et transitions qui décrivent le déroulement séquentiel du processus étudié.

- Des systèmes multi-agents : les réseaux de Petri peuvent modéliser plusieurs agents agissant au sein d'un même système. Nous avons montré dans le cas des wish lists que nous pouvons modéliser le comportement des agents A mais aussi celui des agents B dans un même modèle. Dans le jeu viral, nous avons modélisé le comportement d'un nombre important d'agents (plus de 200 000).

- De nombreuses données : l'une des difficultés de la modélisation de phénomènes sur Internet est la profusion de données disponibles sur plusieurs milliers d'internautes. Nous avons déjà souligné le fait que cet écueil est souvent cité dans les articles de recherche en marketing traitant des problématiques sur Internet. Les réseaux de Petri peuvent gérer cette complexité en intégrant différents types de lois et différents modèles permettant de transformer le nombre important de données en sous-systèmes à analyser. Ils peuvent également gérer de nombreuses données, comme nous avons pu le montrer dans le cas du jeu viral.

- Plus généralement, les réseaux de Petri ont fourni toutes les premières approches de modélisation, ainsi que leurs sémantiques, utilisées pour modéliser et maîtriser les comportements des systèmes parallèles et distribués, synchronisés et communicants (Haas, 2002).

- Ils constituent des représentations faciles à comprendre et donc à manipuler, à la fois pour la création des modèles et pour leur analyse.

- Ils permettent d'exprimer très simplement les concepts premiers des fonctionnements communicants, en incluant les phénomènes d'attente et de synchronisation, et en prenant en considération leurs caractéristiques et paramètres temporels et stochastiques.

- Etant non liés à un langage particulier de réalisation, ils assurent l'indépendance de la modélisation vis-à-vis des cas applications sur lesquels on l'utilise. De plus, les réseaux de Petri sont accessibles à de potentiels utilisateurs de par le fait que plusieurs logiciels d'utilisation des réseaux de Petri sont téléchargeables simplement et gratuitement sur le Web.

- Les réseaux de Petri possèdent un grand intérêt dans de très nombreux et importants domaines d'applications. Ils ont notamment été utilisés dans de nombreuses recherches, en particulier dans les domaines suivants : recherche opérationnelle, informatique, systèmes d'information mais aussi science des organisations. Un site Web (www.daimi.au.dk/PetriNets) particulièrement complet sur les réseaux de Petri regroupe une communauté virtuelle de l'ensemble des chercheurs et praticiens qui dans le monde travaillent sur ce thème.

Les réseaux de Petri peuvent être utilisés pour modéliser et simuler des systèmes de tout type. Ils sont en particulier très utiles pour décrire et analyser des systèmes dynamiques complexes qui comportent des flux d'objets ou d'informations.

L'une des familles les plus riches des réseaux de Petri est la famille des réseaux de Petri stochastiques colorés qui ont toutes les caractéristiques des réseaux de Petri

stochastiques mais qui permettent en plus de distinguer les jetons du réseau. Ceci est particulièrement intéressant pour des applications en marketing pour lesquelles il est nécessaire de distinguer différents types de consommateurs ou différents types de produits.

Les implications managériales :

Les implications managériales issues du travail de cette thèse ont été détaillées dans les différentes applications que nous avons menées. Nous allons donc dans cette conclusion reprendre les points essentiels.
Dans notre première application, le modèle des réseaux de Petri appliqué aux wish lists nous a permis de tirer plusieurs implications pour les managers en marketing :

- Nous avons construit grâce aux réseaux de Petri un modèle permettant de mesurer la performance des wish lists en terme de connexions sur le site et de nombre d'items achetés. Cet outil permet aux gestionnaires de wish lists (millemercis ou des sites comme www.amazon.com) de mesurer et de comparer la performance des différentes wish lists gérées afin de mieux gérer les listes les moins performantes et optimiser le système global.

- Le premier résultat de la mesure de l'impact de l'e-mailing est que tous les internautes ne réagissent pas de la même manière à l'intervalle entre 2 e-mails envoyés. Deux groupes se distinguent : un groupe avec une réaction en U inversé dévoilant un pic de taux d'ouverture de l'e-mail pour un intervalle « optimum » de 18,9 jours; un deuxième groupe avec une réaction en U pour lequel, jusqu'à un intervalle de 28 jours, plus l'intervalle est grand et plus le taux d'ouverture de l'e-mail est élevé, ce taux d'ouverture remontant au-delà de 28 jours.

- Le deuxième résultat lié à l'impact de l'e-mailing est qu'une stratégie d'envoi d'e-mails tous les 7 jours a plus d'impact sur le nombre de connexions et le nombre d'items achetés que la stratégie d'envoi d'e-mails tous les 23 jours (hausse de 5% des achats et progression du nombre de connexions dans les deux groupes).

- Plus généralement, nous proposons par les réseaux de Petri un système d'aide à la décision pour optimiser les campagnes e-mails. Ceci permet notamment aux managers de simuler plusieurs types de campagnes e-mailing en choisissant l'intervalle de temps entre deux e-mails envoyés (mais on pourrait ajouter d'autres critères), d'en mesurer la performance afin de sélectionner la meilleure campagne à lancer.

Dans notre deuxième application, les implications managériales qui ont été présentées en fin du chapitre 6 sont ici reprises. Elles concernent les résultats issus des différentes stratégies simulées et leur impact sur la performance globale du jeu :

- La meilleure stratégie est de doubler l'affiliation, ce qui permet d'accroître de 65% le nombre d'adresses virales fournies et le nombre d'adresses virales converties en joueurs.

- La 2ème stratégie la plus performante est de modifier l'ordre et le type de questions et de proposer une question virale suivie d'une question de qualification plutôt que trois questions de qualification suivies d'une question virale (jeu réel).On augmente dans ce cas la performance du jeu de 50%.

- La 3ème stratégie en terme d'amélioration de performance est l'alternance 2 questions de qualification/ 2 questions virales (+40% d'adresses virales sur 30 jours).

- Viennent ensuite les stratégies de doublement ou triplement de l'e-mail de relance. On constate que ces deux stratégies permettent d'accroître le nombre d'adresses virales de 29%. La stratégie de 2 mails de relance est aussi performante sur 20 jours de jeu que le doublement de l'affiliation qui est coûteux. Cette stratégie est donc à développer, puisque l'envoi d'e-mails représente un coût supplémentaire très faible pour l'entreprise.

- Le fait de doubler le recrutement par autopromotion sur le site Anyway n'améliore pas fortement la performance globale du jeu.

- En diminuant l'affiliation (démarrage uniquement le 15ème jour) et en envoyant 2 mails de relance, on diminue la performance par rapport au jeu réel.

- En terme d'ordre et de choix de type de questions, on constate que la meilleure stratégie (parmi celles testées) est un questionnaire avec 5 questions virales/ 5 questions de qualification, qui permet de multiplier par 2.8 le nombre d'adresses virales fournies. Cette stratégie est préférable à l'ordre inverse où on commence le questionnaire avec 5 questions de qualification suivies de 5 questions virales. Cependant, cette stratégie est à relativiser car on peut se demander si les paramètres du modèle proposé sont encore valides pour ce cas « extrême ».

- Plus globalement, les réseaux de Petri dans cette deuxième application constituent un outil de mesure de la performance d'une opération de marketing viral qui répond au besoin actuel des professionnels en marketing qui est la mesure de la performance d'actions sur Internet.

- Enfin, nous proposons ici un système d'aide à la décision permettant aux marketers de simuler différentes stratégies possibles en terme de marketing viral et en terme de design de questionnaire sur le Web. En effectuant différentes simulations, les professionnels peuvent ainsi tester et mesurer différents scenari avant de sélectionner le plus performant et de l'opérationnaliser en situation réelle. Une application possible de ce système est l'optimisation des questionnaires dans les enquêtes en ligne.

Plus globalement, on peut également souligner que les réseaux de Petri, de par leur représentation graphique simple, peuvent plus facilement être introduits auprès des professionnels que des techniques de modélisation plus complexes. Les modèles de Petri permettant d'étudier des systèmes dynamiques complexes en réduisant leur complexité, ils permettent par conséquent une meilleure compréhension des phénomènes complexes par les professionnels.

Limites des approches développées dans la thèse :

Nous avons montré ci-dessus les avantages des réseaux de Petri et par conséquent l'intérêt du thème développé dans cette thèse. En revanche, les approches que nous avons choisies dans certaines étapes du travail que nous avons mené présentent des limites que nous développons ci-dessous.

- Nous avons été particulièrement vigilante dans les deux applications que nous avons modélisées à ce que chacun des « sous-modèles » décrivant le système global soit un modèle représentant correctement les données observées. Dans le cas des wish lists, nous avons comparé la performance globale du réseau de Petri au modèle NBD avec régression classiquement utilisé dans le domaine du marketing. Nous avons montré dans ce cas que le modèle de Petri donnait de meilleurs résultats. En revanche, nous n'avons pas comparé dans le cas du jeu viral le modèle de Petri à une autre méthode de modélisation de ce type de système. Ceci constitue une première limite de notre travail mais constitue également une piste de recherche future,

- Notre première application, la modélisation des wish lists, comporte un certain nombre de limites que nous avons déjà décrites dans le chapitre 4 mais que nous reprenons et développons ici :
    - Nous avons comparé un modèle individuel (les paramètres des réseaux de Petri sont différents pour chaque agent, tenant compte ainsi de l'hétérogénéité des agents) à un modèle NBD avec régression dont les paramètres globaux sont les mêmes pour chaque individu, même si ces paramètres ont été calculés à partir de données individuelles. Il eût été plus pertinent de comparer des modèles homogènes sur ce critère.

    - La deuxième limite que nous avons déjà soulignée est le nombre élevé de paramètres estimés par les réseaux de Petri (6 paramètres pour chacun des 94 agents). Nous avons déjà tenté d'améliorer ce point et montré dans le chapitre 3 que les comportements de connexion des agents A et B comme leur comportement d'achat pouvaient être modélisés par des modèles agrégés sur l'ensemble des agents (nous avons par exemple pour cela étudié les lois suivies par chaque paramètre de la loi de Weibull représentant les connexion des agents). En revanche, nous n'avons pas pu paramétrer dans le logiciel SPNP des réseaux de

Petri les fonctions Weibull ou Gamma qui définissent les paramètres de la loi de Weibull de la transition de connexion des agents (dans le logiciel, les transitions peuvent utiliser des lois de Weibull par exemple, mais pas les paramètres des lois des transitions). La limite que nous évoquons ici est donc liée à une limite technique actuelle du logiciel SPNP utilisé dans cette thèse,

- Nous utilisons pour la durée entre deux connexions des lois de Weibull qui sont stationnaires. Certains chercheurs (comme Moe et Fader par exemple) ont développé des modèles non stationnaires (exponentiel non stationnaire) pour tenir compte de l'expérience de l'internaute sur le site. Cette approche pourrait être intéressante à intégrer dans notre modèle.

- Nous avons sélectionné dans cette application des wish lists particulièrement actives. Il aurait également été intéressant d'appliquer le modèle à des utilisateurs plus modérés de wish list. Nous avons ensuite pour des raisons de calibration du modèle sélectionné 94 agents sur lesquels 16 mois de données étaient disponibles, ce qui a encore réduit l'échantillon utilisé au final. De plus, certains phénomènes comme la saisonnalité éventuelle des achats ou les périodes de vacances n'ont pas été contrôlées.

- Une autre limite de notre approche est que nous n'avons pas intégré toutes les variables du comportement de l'internaute sur le site millemercis. En particulier, comment les internautes naviguent-ils sur le site ? Quelle est la durée moyenne des visites ? Quelles sont les pages vues ? Nous aurions pu également approfondir le type de listes et le type de produits par liste (incluant la notion de prix). Ces caractéristiques enrichiraient le modèle que nous avons développé. Nous ne les avons pas volontairement incluses dans le modèle de base, notre objectif dans la thèse étant de présenter une première application simple des réseaux de Petri à une problématique de marketing, nous voulions traiter un cas simple de manière à bien montrer le principe et l'intérêt de la méthode.

- En ce qui concerne la mesure de l'impact de l'e-mailing, il est clair que le partage des agents sur les groupes 1 et 2 permet d'introduire une certaine hétérogénéité

dans le modèle, mais cette méthode est plutôt ad hoc. D'autres méthodes de segmentation pourraient être appliquées.

- Nous avons utilisé pour modéliser les campagnes e-mailing une base de données d'e-mails effectués par une seule entreprise (millemercis, même si ces e-mails étaient pour le compte de différentes entreprises). Il serait intéressant d'utiliser d'autres bases de données pour généraliser les résultats.

- Notre deuxième application, le jeu viral promotionnel, présente également un certain nombre de limites :
  - Nous avons modélisé l'opération « le ciel est à tout le monde » par les réseaux de Petri sans comparaison avec d'autres approches plus classiques en marketing (comme les chaînes de Markov par exemple) afin de comparer les performances du modèle de Petri avec ces autres méthodes.

  - Le modèle de transformation des adresses virales en joueurs n'intègre pas le temps. Il serait intéressant d'étudier si le taux de conversion est le même en début de jeu et en fin de jeu.

  - Plusieurs éléments de l'effet viral du jeu restent à approfondir. Tout d'abord, il serait intéressant d'étudier l'effet de « cercle » : au bout de combien d'adresses virales fournies revient-on sur une adresse déjà fournie ou qui a déjà joué ? D'autre part, quel est le profil des bons « viralisateurs » (ceux qui fournissent un nombre important d'adresses virales) par rapport aux moins bons ?

  - Le modèle de Petri nous a permis de mesurer la performance globale du jeu sur le critère du nombre d'adresses virales fournies et du nombre d'adresses virales joueuses. D'autres critères pourraient être intégrés comme par exemple le nombre de bons « viralisateurs ».

  - Une autre limite de notre approche est le fait que nous avons travaillé sur une seul jeu viral. Pour généraliser les résultats et valider les modèles utilisés, il serait intéressant d'appliquer les réseaux de Petri sur d'autres bases de données de ce type de jeu.

Pistes de recherche futures

Le thème même de la thèse est source de développement de pistes futures de recherche. En effet, l'objectif de la thèse étant d'introduire les réseaux de Petri pour modéliser des phénomènes marketing sur Internet, de nombreuses pistes s'ouvrent dans cette optique. Certains de ces thèmes de recherche sont liés aux limites que nous avons développées ci-dessus. Nous exposons ci-dessous l'ensemble des pistes possibles :

- La première piste de recherche est de comparer la performance des réseaux de Petri à plusieurs modèles déjà utilisés en marketing. Sur Internet en particulier, on peut penser aux travaux récents de Moe et Fader (2004) portant sur un modèle de conversion des visites en achats qui est devenu une référence. Ce modèle constitue un modèle de comparaison intéressant. L'objectif de cette piste de recherche serait donc d'appliquer sur une même base de données le modèle de Moe et Fader et le modèle des réseaux de Petri afin de comparer la performance de ces deux approches. On peut également utiliser plus directement le travail de la thèse et tenter de modéliser le jeu viral par une autre méthode (par exemple des chaînes de Markov) pour montrer l'éventuelle meilleure performance des réseaux de Petri.

- Sur la modélisation du système de wish list, plusieurs pistes de recherche sont ouvertes :
    - L'une de ces pistes est de développer un modèle de comparaison tenant compte de l'hétérogénéité des agents et directement comparable sur ce critère au modèle des réseaux de Petri utilisé.

    - La deuxième piste possible est d'explorer l'ensemble des logiciels des réseaux de Petri stochastiques afin de déterminer si certains de ces logiciels permettent d'intégrer des lois de type Weibull et Gamma applicables aux paramètres des lois de Weibull des transitions. L'objectif ici est de proposer un modèle de Petri agrégé au niveau de l'ensemble des individus, ce qui diminuerait fortement le nombre de paramètres du modèle. Le logiciel que nous avons utilisé n'intègre pas actuellement cette possibilité. Il serait donc intéressant de comparer de manière exhaustive les différents logiciels existant pour représenter les réseaux de Petri. On peut aussi envisager un développement spécifique sur ces logiciels.

- Il serait également intéressant d'introduire pour la modélisation des connexions des agents A et B des lois non stationnaires, comme celles développées par Moe et Fader (2004).

- La quatrième piste consiste à développer le modèle de Petri sur un échantillon beaucoup plus important de wish lists, comportant des agents actifs mais aussi des utilisateurs plus modérés, voire très modérés, de ce type de système, afin d'avoir des résultats plus généralisables.

- La cinquième piste est une recherche sur l'intégration d'autres variables de comportement dans le modèle : comportement de navigation, pages vues, durée de visite, type de liste, prix des produits, ce qui permettrait de complexifier le système tout en fournissant en contrepartie un modèle plus explicatif du comportement des internautes sur les wish lists.

- La sixième piste concerne l'impact de l'e-mailing. Il serait intéressant de développer d'autres approches de segmentation permettant de définir plusieurs groupes de réponse aux campagnes e-mailing.

- La septième piste consiste à appliquer le modèle d'impact de l'e-mailing à d'autres bases de données de campagnes d'e-mails, notamment des campagnes menées par des entreprises différentes, afin d'accroître la généralité des conclusions (notamment l'existence de plusieurs groupes de réponse à l'e-mail et l'intérêt de développer des campagnes e-mail tous les 7 jours plutôt que tous les 23 jours).

- En ce qui concerne la modélisation du jeu viral promotionnel on line, les pistes de recherche futures sont les suivantes :
  - Il serait intéressant de développer des modèles de comparaison (nous avons déjà développé ce point) afin de bien mesurer l'apport des réseaux de Petri par rapport à des modèles plus traditionnels en marketing.

  - Le modèle de conversion des adresses virales en joueurs pourrait être amélioré en faisant varier les paramètres du modèle log-réciproque dans le temps.

- La troisième piste de recherche sur le jeu promotionnel concerne l'étude de l'effet de cercle ci-dessus décrit. L'idée ici serait d'introduire cet effet dans le réseau de Petri afin de déterminer s'il existe un temps fini au bout duquel plus aucune adresse virale supplémentaire n'est générée.

- La quatrième piste de recherche concerne la segmentation des joueurs en « bons viralisateurs » ou « moins bons viralisateurs ». Cette segmentation pourrait être menée en fonction de critères comme le nombre d'adresses virales fournies et transformées par individu mais également les profils (profils socio-démographiques classiques ou profils de comportement vis-à-vis du voyage fournis par les réponses aux questions de qualification du jeu). Il serait également intéressant de relier cette segmentation à des critères de type Customer Lifetime Value (est-ce-que les « bons viralisateurs » sont de bons clients par exemple ?)

- La cinquième piste de recherche est l'application du modèle de Petri à d'autres jeux de marketing viral afin de valider le modèle sur d'autres applications et d'étendre la validité externe du modèle et la généralité des résultats.

- La sixième piste de recherche possible est l'utilisation des réseaux de Petri stochastiques colorés intégrant tous les éléments de notre modèle mais différenciant des types d'agents (cette recherche se situe après la segmentation sur le critère « bons viralisateurs » par exemple). Ceci permettrait notamment d'effectuer des simulations incluant ce critère (que se passe-t-il lorsqu'on augmente le nombre de « bons viralisateurs » dans le recrutement ?

- Plus généralement, l'application des réseaux de Pétri ouvre de nombreuses pistes de recherche en marketing, ce type de modélisation n'ayant quasiment pas été utilisé dans ce domaine. Plusieurs problématiques marketing peuvent être modélisées par cette méthode : Nous distinguerons des problématiques marketing « classiques » et des problématiques liées aux nouvelles technologies, pour lesquelles les réseaux de Petri sont particulièrement adaptés.

Dans le premier cas, les réseaux de Petri pourraient être utilisés pour modéliser par exemple :

- Les services après-vente et la gestion des réclamations,
- la décomposition des étapes du processus d'achat d'un consommateur
- l'optimisation du processus de développement d'un nouveau produit, en particulier le processus de travail collaboratif mené dans ce contexte,
- l'optimisation de l'organisation des équipes marketing,
- de nombreuses applications en B to B, dans lequel l'optimisation de systèmes de réseaux est une problématique importante,
- l'optimisation des rayons des distributeurs, et de nombreuses autres applications pour la grande distribution (performance des supply chains, réaction des consommateurs aux prix, processus de négociation avec les fabricants, gestion de files d'attentes aux caisses, etc..).

En ce qui concerne les thèmes marketing liés aux nouvelles technologies, les applications des réseaux de Petri pourraient concerner :

- l'étude du comportement de l'internaute sur un site Web (modélisation du comportement de visite et d'achat),
- l'optimisation de la conception des interfaces et la gestion des sites Web,
- la modélisation des échanges au sein des communautés virtuelles,
- la communication par téléphone mobile entre plusieurs agents,
- l'optimisation d'opérations de marketing viral,
- la modélisation des enchères en ligne,
- l'optimisation des questionnaires dans les enquêtes en ligne,
- plus globalement toutes les problématiques marketing liées à des phénomènes de réseaux.

En conclusion, nous souhaitons que les éléments présentés dans cette thèse suscitent et motivent des chercheurs en marketing pour de futurs travaux permettant en particulier de mieux comprendre le consommateur à l'ère complexe des nouvelles technologies, que ce soit

Internet, la téléphonie mobile ou demain la télévision interactive. Nous espérons ainsi avoir contribué, à l'échelle de cette thèse, à une meilleure compréhension du monde de demain.

Bibliographie utilisée dans la thèse:

Ansari, A., & Mela, C.F. (2003). E-customization. *Journal of Marketing Research*, *40*(2), 131-145.

Armstrong, A., & Hagel, J. III (1996). The real value of on-line communities. *Harvard Business Review*, May-June.

Arndt, J. (1967a). Role of product-related conversations in the diffusion of a new product. *Journal of Marketing Research*, 4, 291-95.

Arndt, J. (1967b). *Word of mouth advertising: a review of litterature.* New york: Advertising research Federation.

Arndt, J. (1968). Selective process in word-of-mouth. *Journal of Advertising Research*, 8, 19-22.

Bagozzi, R.P., & Dholakia, U.M. (2002). Intentional social action in virtual communities. *Journal of Interactive Marketing*. Vol.16, No2.

Bakam, Kordon, Le Page, Bousquet (1999) "Formalisation de modèles multi-agents par les réseaux de Petri: Application à l'étude d'un système de gestion de la chasse à l'Est-Cameroun" , *Centre De Coopération Internationale en Recherche Agronomique pour le Développement.*

Bakos, J.Y. (1997). Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science*. Vol.23, No12.

Bansal, H.S. et Voyer, P.A. (2000). Word-of-mouth processes within a services purchase decision context. *Journal of Service Research*, Vol.3, No2, 166-177.

Bates, J., Bryan Loyall, A., and Scott Reilly, W. (1992). Integrating reactivity, goals and emotion in a broad agent. *Technical report CMU-CS-92-142. School of Computer Science*, Carnegie-Mellon University, Pittsburgh, PA.

Bates, J., & Bryan Loyall, A., & Scott Reilly, W. (1992). An architecture for action, emotion and social behaviour. *Technical report CMU-CS-92-144. School of Computer Science*, Carnegie-Mellon University, Pittsburgh, PA.

Bemmaor, Albert C. (1984). Testing alternative econometric models on the existence of advertising threshold effect. *Journal of Marketing research*, Vol. 21, pp. 298-308.

Betts, M. (2001). Turning browsers into buyers. *MIT Sloan Management Review, 42*(2), 8-9.

Brodin, O. (2000). Les communautsé virtuelles: un potentiel marketing encore peu explore. *Décisions Marketing*, No 21. Sept-Déc.

Bowman, D. et Narayandas, D. (2001). Managing customer-initiated contacts with manufacturers : the impact on share of category requirements and word-of-mouth behavior. *Journal of Marketing research*, vol. 38, 281-297.

Bristor, J.M. (1990). Enhanced explanations of word of mouth communications: the power of relationships.*Research in Consumer Behavior*, 4, 51-83.

Brown, J.J. et Reingen, P.H. (1987). Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, 14, 350-62.

Brown, W.M. et Tucker, W.T. (1961). Vanishing shelf space. *Atlantic economic review*, Vol.9, pp9-13.

Bucklin, R.E., & Sismeiro, C. (2003). A model of Web site browsing behavior estimated on clickstream data, *Journal of Marketing Research*, *40*(3), 249-267.

Bughin, J., & Hagel III, J. (2000). The Operational Performance of Virtual Communities-Towards a Successful Business Model?. *Electronic Markets*, Vol. 10, Issue 4.

Berthon, P., & Leyland F., P. (1996). The World Wide Web as an Advertising Medium: Toward an Understanding of Conversion Efficiency. *Journal of Advertising Research*, Vol.36 Issue 1, 43-55.

Berzjian-Avery, Calder and Iacobucci (1998). New media interactive advertising vs.traditional advertising. *Journal of Advertising Research*, Jul/Aug 1998.

Betts, M. (2001). Turning browsers into buyers. *MIT Sloan Management Review, 42*(2), 8-9.

Bickart, B., & Schindler, R.M. (2001). Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*. Vol. 15, No3.

Blattberg, R.C., & Deighton, J. (1991). Interactive Marketing: Exploiting the Age of Adressability, *Sloan Management Review*, Fall 1991.

Burgoon, Bonito, Bengtsson, Ramirez (Winter 1999/2000), "Testing interactivity model: Communication processes, partner assessments, and the quality of collaborative work", *Journal of Management Information Systems.*

Card, S.K., & Moran, T.P., & Newell, A. (1980). The Keystroke-Level Model for User Performance Time with Interactive Systems. *Communications of the ACM*, Vol. 23 Issue 7, 396-411.

Celsi, R.L., & Randall, L.R., Leigh, T.W. (1993). An exploration of high-risk leisure consumption through skydiving. *Journal of Consumer Research*, Vol.20.

Chavez , A. and Maes, P. (1996) Kasbah: an agent marketplace for buying and selling goods. In *Proceedings of the 1rst International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology* (PAAM-96), London, UK, pp.75-90.

Cho,C., Chang-Hoan, & Leckenby, J.D. (1997). Internet-related Programming Technology and Advertising. In *Proceedings of the 1997 Conference of the American Academy of Advertising,* M.Carole Macklin, ed., Cincinnati, OH: American Academy of Advertising, 69.

Cho, C., & Leckenby, J.D. (1999). Interactivity as a Measure of Advertising Effectiveness: Antecedents and Consequences of Interactivity in Web Advertising. In *Proceedings of the 1999 Conference of the American Academy of Advertising*, M.S. Roberts, ed. Gainesville, FL: University of Florida.

Conte, R. & Gilbert, N. (1995). Computer simulation for social theory. In *Artificial Societies: The Computer Simulation of Social Life* (eds N. Gilbert and R. Conte), pp. 1-15. UCL Press, London.

Cova, B. (1996). Leçons de marketing "postmoderne", *L'expansion Management Review*, Décembre, 91-101.

Cova, B., & Roncaglio, M. (1999). Repérer et soutenir des tribus de consommateurs, *Décisions Marketing*, No 16, Janvier/Avril, 7-15.

Cova, B. & Cova, V. (2002). Tribal Marketing: The Tribalization of Society and Its Impacts on the Conduct of Marketing. *European Journal of Marketing*, Vol. 36, N°5/6, p. 595-620.

Cova, V. & Cova B. (2001). *Alternatives Marketing* : *réponses marketing aux nouveaux consommateurs*, Dunod. Paris.

Coviello, N., Milley, R., & Marcolin, B. (2001). Understanding IT-enabled interactivity in contemporary marketing. *Journal of Interactive Marketing*, Vol. 15, No 4.

Coyle, J.R., & Thorson, E. (2001). The effects of Progressive levels of Interactivity and Vividness in Web Marketing Sites. *Journal of Advertising*, vol. 30, (3) 65-77.

Csikszentmihalyi, M. (1977*). Beyond Boredom and Anxiety*. San Fransisco: Jossey-Bass.

Csikszentmihalyi, M. (1983). Measuring Intrinsic Motivation in everyday Life. *Leisure studies,* 2, 155-168.

Csikszentmihalyi, M. (1988). The future of flow, in *Optimal Experience: Psychological Studies of Flow in Consciousness*. Mihaly Csikszentmihalyi and Isabella Selega Csikszentmihalyi, eds. Cambridge: Cambridge University Press. 364-383.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.

Csikszentmihalyi, M., & Csikszentmihalyi, I.S. (1988). *Optimal Experience: Psychological Studies of Flow in Consciousness*. Cambridge: Cambridge University Press.

Csikszentmihalyi, M., & Lefebvre, J. (1989). Optimal Experience in Work and Leisure. *Journal of Personality and Social Psychology*, 56 (5), 815-822.

Cuneo, A.Z. (1995). Internet World Show spurs online commerce debate. *Advertising Age*, Vol.66 Issue 16, 16-21.

Darpy, D., & Volle, P. (1999). *Connaissance des consommateurs et décisions marketing*. Polycopiés cours de comportement du consommateur, Université Paris Dauphine, 1999-2000.

David rené and Alla Hassine, "Du Grafcet aux réseaux de Petri », *Editions Hermès, 1997.*

Deighton, J. (1996). The Future of Interactive Marketing. *Havard Business Review*, November-December.

Dholakia, U.M., & Bagozzi, R.P., & Kelin Pearo, L. (2004). A social inluence model of consumer participation in network- and small group-based virtual communities. *International Journal of Research in Marketing*, Vol. 21, Issue 3. 241-263.

Diaz Michel, "Les réseaux de Petri, modèles fondamentaux", *Editions Hermès, 2001, Lavoisier.*

Diaz Michel, "Vérification et mise en œuvre des réseaux de Petri", *Editions Hermès, 2003, Lavoisier.*

Doise W., Deschamps, J.C., Mugny G. (1991). *Psychologie Sociale Expériementale*. Armand Colin.

Doorenbos, R., Etzioni, O. and Weld D. (1997) A scaleable comparison-shopping agent for the world wide web. In *Proceedings of the International Conference on Autonomous Agents* (Agents 97), Marina del Rey, CA pp.39-48.

Doran, J. (1987). Distributed artificial intelligence and the modeling of socio-cultural systems. In *Intelligent Systems in a Human Context* (eds L.A. Murray and J.T. E. Ricahrdson), pp. 71-91. Oxford University Press, Oxford.

Doran, J., & Palmer, M. (1995). The EOS project: integrating two models of Paleolithic social change. In *Artificial Societies: the Computer Simulation of Social Life* (eds N. Gilbert and R. Conte), pp. 103-125? UCL Press, London.

Doran, J., et al. (1992). The EOS project . In *Simulating Societies-92: Pre-proceedings of the 1992 International Symposium on Approaches to Simulating Social Phenomena and Social Processes, Department of Sociology, University of Surrey.*

Drèze, X., & Zufryden, F. (2004). Measurement of online visibility and its impact on internet traffic. *Journal of Interactive Marketing*. Vol. 18, No 1.

Duncan, T., & Moriarty, S.S (1998). A Communication-Based Marketing Model for Managing relationships. *Journal of Marketing*, 62, April, 1-13.

Dye, R. (2000). The buzz on buzz. *Harvard Business Review*, 78 (6), 139-147.

Evrard, Y., & Pras, B., & Roux, E. (2000*). Market, Etudes et recherches en Marketing.* Dunod. Paris.

Etzioni, O., & Weld, D.S. (1995). Intelligent agents on the Internet: fact, fiction and forecast. *IEEE expert*, 10(4), 44-49.

Etzioni, O. (1996). Moving up the information food chain: deploying softbots on the World Wide Web. In *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI-96), Portland, OR*, pp.4-8.

Ferber, J. (1995). *Les systèmes multi-agents*, *vers une intelligence collective.* InterEditions.

Ferber (1997). Les systèmes multi-agents : un aperçu. Dans: *Technique et science informatiques.* Vol.16, No 8, 979-1012.

Fischer, G.N. (1997). *La Psychologie Sociale*, Editions du Seuil, Essais.

Fischer, E., & Bristor, J., & Gainer, B. (1996). Creating or escaping community ?: An exploratory study of Internet consumers' behaviors. *Advances in consumer research*, Vol.23.

Ghose, S., & Dou, W. (1998). Interactive functions and their impacts on the appeal of internet presence sites, *Journal of Advertising Research,* March/April.

Godin, S. (2001). *Les secrets du marketing viral*. Editions Maxima, Paris.

Gorn, G.J., & Chattopadhyay, A, & Sengupta, J;& Tripathi, S. (2004). Waiting for the Web: how screen color affects time perception. *Journal of Marketing Research*, Vol. 61, No2, 215-225.

Granitz, N.A., & Ward, J.C. (1996). Virtual community: a Sociocognitive analysis. *Advances in Consumer Research*. Vol. 23.

Gruen, T. (forthcoming). "Integrated Marketing Communications and the Emerging Role of the Web Site". In Allan Kimmel (Vol. Ed.). Marketing Communication: New Approaches, Technologies and Styles.

Gugel, C. (1997). The interactive telemedia index: An Internet/ITV impact model. *Journal of Advertising Research*, Vol. 37, Issue 2, 29-32.

Gupta, U.G. (1997). Using citation analysis to explore the intellectual base, knowledge dissemination, and research impact of Interfaces (1970-1992), *Interfaces,* 27, pp85-101.

Guttman, R.H., & Moukas, A.G., & Maes, P. (1998). Agent-mediated electronic commerce: a survey. *The Kowledge Engineering Review*, 13(2), 147-159.

Ha, L., & James, E.L. (1998). Interactivity reexamined : A baseline analysis of early business Web sites. *Journal of broadcasting & Electronic Media*, Vol.42 Issue 4, 456-474.

Haas, P. J. (2002). Stochastic Petri Nets. In P.W. Glynn & S.M. Robinson (Series Ed.), *Springer Series in Operations Research*. New-York: Springer-Verlag.

Habermas , J. (1981). *Theorie des kommunikativen Handelns*, Suhrkamp, (2 volumes).

Habermas , J. (1990). *Moral consciousness and communicative action*. Trans Ch. Lenhardt and Sh. W. Nicholsen, MIT Press.

Hagel , J. (1999). Net Gain: Expanding markets through virtual communities. *Journal of Interactive Marketing*, Vol.13, No1.

Hagel III, J., & Armstrong, A.G. (1997). *Net Gain: Expanding Markets Through Virtual Communities*. Harvard Business School Press.

Hagel III, J., & Armstrong, A.G. (1999). Bénéfices sur le Net. Editions d'organisation. (traduction de l'ouvrage original *Net Gain: Expanding Markets Through Virtual Communities*. Harvard Business School Press, 1997).

Hardie et Fader, disponible sur le site http://brucehardie.com/talks.html:

Applied Probability Models in Marketing Research: Introduction
(Tutorial at the AMA's Advanced Research Techniques Forum, Whistler, B.C., June 2004)F

Forecasting Repeat Buying for New Products and Services
(Tutorial at the AMA's Advanced Research Techniques Forum, Whistler, B.C., June 2004)

Applied Probability Models in Marketing: An Introduction
(Statistical Innovations "Statistical Modeling Week" Workshop, October 2002)

Materials for previous ART Forum Tutorials (2000-2003)

http://brucehardie.com/notes/

- Spreadsheet to accompany "A Note on an Integrated Model of Customer Buying Behavior"
- Implementing the BG/NBD Model for Customer Base Analysis in Excel
- Illustrating the Performance of the NBD as a Benchmark Model for Customer-Base Analysis

http://brucehardie.com/papers.html

Fader, Peter S., Bruce G.S. Hardie, and Ka Lok Lee (2004), "'Counting Your Customers' the Easy Way: An Alternative to the Pareto/NBD Model."

Harrisson-Walker, L.J. (2001). The measurement of word-of-mouth communication and an investigation of service quality and customer commitment as potential antecedents. *Journal of Service Research*, Vol. 4, No1, 60-75.

Helm, S. (2000). Viral marketing-establishing customer relationships by "word-of-mouse". *Electronic Markets*, Vol. 10 (3), 158-161.

Hoffman, D.L, & Novak, T.P. (1996). Marketing in hypermedia Computer-Mediated Environments: Conceptual Foundations. *Journal of Marketing*, 60 (3), 50-68.

Hoffman, D.L., Novak, T.P., & Chatterjee, P. (1995). Commercial Scenarios for the Web: Opportunities and Challenge. *Journal of Computer-mediated Communication*, Special Issue on electronic Commerce 1, 3.

Holbrook, M.B., Chestnut, R.W., Oliva, T.A., Greenleaf, E.A. (1984). Play as a Consumption Experience: The Roles of Emotions, Performance, and Personality in the Enjoyment of Games. *Journal of Consume Research*, 11, 728-739.

Jee, J., & Lee, W.N. (2002). Antecedents and Consequences of Perceived Interactivity: An Exploratory Study. *Journal of Interactive Advertising*, Vol.3, No1.

Johnson, E.J., & Bellman, S., & Lohse, G.L. (2003). Cognitive Lock-In and Power Law of Practice. *Journal of Marketing*, Vol. 67, April, 62-75.

Johnson, E.J., & Moe, W.W., & Fader, P.S., & Bellman, S., & Lohse, G.L. (2004). On the depth and dynamics of online search behavior. *Management Science*. Vol. 50, No3, 299-308.

Kaikati, A.M. et Kaikati, J.G. (2004). Stealth marketing : How to reach consumers surreptitiously. *California Management Review*, Vol. 4, No4, 6-22.

Katz, E. & Lazarsfeld P.F. (1955). *Personal influence*. Glencoe, IL : Free Press.

Khermouch, G. (2001). Buzz Marketing. *Business Week*. July 30.

Komito, KL. (1998). The Net as a foraging society: flexible communities. *Information Society,* 14, 97-106.

Kotler, P., & Dubois, B. (2000). *Marketing Management*. 10ème édtition. Publi-Union Editions. Paris. 199-200.

Kozinets, R.V. (1999). E-tribalized marketing ?: the strategic implications of virtual communities of consumption. *European Management Journal*, Vol.17, No3.

Kozinets, R.V. (2002). The field behind the screen: using netnography for marketing research in online communities. *Journal of Marketing Research*, Vol. 39. 61-72.

Kuokka, D.R., & Harada, L.P. (1996). Issues and extensions for information matchmaking protocols. *International Journal of Cooperative Information Systems*, 5 (2-3), 251-274.

Ladwein, R. (1999). *Le Comportement du Consommateur et de l'Acheteur*, Economica.

Lane, M.S., Mansour, A.H. & Harpell, J.L. (1993). Operations Research Techniques : A Longitudinal Update 1973-1988, *Interfaces,* 23, pp 63-68.

Law, A.M., & Kelton, W.D. (2000). Simulation modeling and analysis. In K.E. Case & P.M. Wolfe (Series Ed.), *McGraw-Hill Series in Industrial Engineering and Management Science*, 3d edition. Singapore: McGraw-Hill Higher Education.

Li, S. (2004). Modeling Online Browsing and Path Analysis Using Clickstream Data. *John A. Howard AMA doctoral Dissertation Award.*

Lilien, G.L., Kotler, P., K. Sridhar, M. (1992). *Marketing Models*. Prentice-Hall Inc., Upper Saddle River, NJ 07458, USA.

Lipovetsky, G. *L'ère du vide*(1983). *L'Empire de l'éphémère* (1987). *Le Crépuscule du devoir* (1992). Tous ces ouvrages sont publiés chez Gallimard.

Liu, Y. (2001). Understanding the interactive media: interactivity and its implications for consumer behavior. *Ph.D Dissertation Proposal*, Rutgers University.

Liu, Y. (2003). Developing a Scale to Measure the Interactivity of Websites. *Journal of Advertising Research*, Vol. 43, Issue 2, 207-307.

Liu, Y., & Shrum, L.J. (2002). What Is Interactivity and Is It Always Such a Good Thing? Implications of Definition, Person, and Situation for the Influence of Interactivity on Advertising Effectiveness. *Journal of Advertising*, Vol. 31, No 4.

Lynch, J.G. Jr., & Ariely, D. (2000). Wine online: Search costs affect competition on price, quality, and distribution. *Marketing Science*, *19*(1), 83-103.

Lyotard, J.F. (1979). *La condition post-moderne*. Paris. Editions de Minuit.

Maes, P. (1994). Social interface agents: acquiring competence by learning from users and other agents. On *Software Agents-Papers from the 1994 Spring Symposium* (ef. O. Etzioni), technical report SS-94-03, pp. 71-78. AAAI Press.

Maffesoli, M. *Le Temps des tribus* (1988*) Paris,* Méridiens; *Au creux des apparences* (1990) Paris, Plon ; *La Contemplation du monde* (1993). Paris, Grasset. *Eloge de la raison sensible* (1996). Paris, Grasset.

McMillan, S. J. (2000). Interactivity Is in the Eye of the Beholder: Function, Perception, Involvement, and Attitude toward the Web Site. In M. A. Shaver (Ed.), *Proceedings of the*

2000 Conference of the American Academy of Advertising (pp. 71-78). East Lansing, MI: Michigan State University.+

McMillan, S.J., & Hwang, J.S. (2002). Measures of Perceived Interactivity: An Exploration of the Role of Direction of Communication, User Control, and Time in Shaping ërceptions of Interactivity. *Journal of Advertising*, Vol. 31, No 3.

Mandel, N., & Jonhson, E.J. (2002). When Web pages influence choice: Effects of visual primes on experts and novices. *Journal of Consumer Research*, *29*(2), 235-245.

Mandiau, R., & Grislin-Le Strugeon, E., & Péninou, A. (2002). *Organisation et applications des SMA*. Editions Hermès. Lavoisier. Paris.

Marney, Jo (1995). Selling in Tongues. *Marketing Magazine*, 100 (38), 14.

Meldman, J.A. (1977). A New technique for Modeling the Behavior of Man-Machine Information Systems. *Sloan Management Review*, *18*(3), 29-47.

Moe, W.W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology*, *13*(1&2), 29-40.

Moe,W.W., & Fader, P.S. (2004a). Capturing evolving visit behavior in clickstream data. *Journal of Interactive Marketing*, *18*(1), 5-19.

Moe, W.W., & Fader, P.S. (2004b). Dynamic conversion behavior at e-commerce sites. *Management Science*, *50*(3), 326-335

Montgomery, A.L., & Li, S., & Srinivasan, K, & Liechty, J.C. (2004). Modeling Online Browsing and Path Analysis using clickstream data. *Marketing Science*. Forthcoming.

Moore, E.G., & Mazvancheryl, S.K., & Rego, L.L. (1996). The Bolo Game: Exploration of a High-Tech Virtual Community. *Advances in Consumer Research*. Vol. 23.

Morrison, D.G., & Schmittlein, D.C. (1988). Generalizing the NBD Model for customer purchases: what are the implications and is it worth the effort?. Journal of Business and Economic Statistics, Vol. 6, no 2 (April), pp. 145-159.

Moscovici, S. (1984). *Psychologie sociale*. Presses Universitaires de France.

Muniz, JR., A.M., & O'Guinn, T.C. (2001). Brand Community. *Journal of Consumer Research*. Vol. 27, Issue 4.

Negroponte, N. (1995). *Being Digital*. Hodder and Stoughton, London.

Noriega, P., & Sierra, C. (eds) (1999). *Agent Mediated Electronic Commerce*. LNAI Volume 1571. Spinger, Berlin.

Normann, R. & Ramirez, R. (1993). From value chain to value constellation : designing interactive strategy. *Havard Business Review*, Vol.71, Issue 4, 65-77.

Normann, R. & Ramirez, R. (1994). *Designing Interactive Strategy*. John Wiley & Sons. 2e éd.. Traduction en japonais Sannon Daigaku, 1996.

Novak, T.P., & Hoffman, D.L., & Yung, Y.F.. Measuring the customer experience in online environments : a structural modeling approach. *Marketing Science*, Vol.19, No 1, 22-42.

Papazoglou, M.P., & Laufman, S.C., & Sellis, T.K. (1992). An organizational framework for cooperating intelligent information systems. *Journal of Intelligent and Cooperative Information Systems*, 1(1), 169-202.

Park, Y.H., & Fader, P.S. (2004). Modeling browsing behavior at multiple websites. *Marketing science*, Vol. 23, No3, 280-303.

Petri, C.A. (1962). Kommunikation mit automaten. PhD thesis, Shriften des Institutes fur Instrumentelle Mathematik Bonn, Germany (in German).

Petri Nets World: http://www.daimi.au.dk/PetriNets.

Rafaeli, S. (1988). Interactivity: From new media to communication. In R.P. Hawkins, J.M. Reeves Byron, Nass Clifford (1996). The Media Equation, how people treat computers, television, and new media like real people and places. *Cambridge university Press.*

Raman, N.V., & Leckenby, J.D. (1998). Factors affecting consumers "Webad" visits. *European Journal of Marketing*. Vol.32, No 7/8, 737-748.

Ramirez, R. (1999). Value co-production: intellectual origins and implications for practice and research. *Strategic management Journal*. Vol. 20. 49-65.

Rayport, J.F., & Jaworski, B.J. (2002). Introduction to e-commerce. New York: McGraw-Hill.

Rheingold, H (1993). *The Virtual Community*. Reading MA: Addison-Wesley.

Salimifard, K., & Wright, M. (2001). Petri net-based modelling of workflow systems: An overview. *European Journal of Operational Research*, *134* (3), 664-676.

Schmittlein, D.C., Bemmaor, A.C., Morrison, D.G. (1985). Why does the NBD model work? Robutsness in representing product purchases, brand purchases and imperfectly recorded purchases. *Marketing Science*, Vol. 4 (3), 255-266.

Shankar, V., Smith, K., & Rangaswamy, A. (2000). The relationship between Customer Satisfaction and Loyalty in Online and Offline Environements. *Pennsylvania State University eBusiness Research center Working paper 02-2000.*

Silverman, G. (1997). Harvesting the power of word of mouth. *Potentials in Marketing,* 30 (9), 14-16.

Sismeiro, C., & Bucklin, R.E. (2004). Modeling purchase behavior at an e-commerce Web site: A task completion approach. *Journal of Marketing Research*, *41*(3), 306-323.

SPNP User's Manual, version 6.0. Available from http://www.ee.duke.edu/~kst/ software

packages.
Stambouli K.B., Briones, E. (2002). *Buzz marketing, les stratégies du bouche-à-oreille.* Editions d'organisation, Paris.
Steuer, J. (1992). Defining Virtual Reality: Dimensions Determining Telepresence, *Journal of Communication,* 42 (4), Autumn.
Tax, S.S., Chandrashekaran, M., Christiansen, T. (1993). Word-of-mouth in consumer decision-making: an agenda for research. *Journal of Customer Satisfaction, Dissatisfaction & Complaining Behavior*, 6, 75-80.
Telang, R., & Boatwright, P., & Mukhopadhyay, T. (2004). A mixture model for internet search-engine visits. *Journal of Marketing Research*, Vol. 61, No2, 206-214.
Trivedi, K.S. (2002). *Probability and statistics with reliability, queuing and computer science applications*. Second Edition. John Wiley & Sons, Inc., New York.
Tsalgatidou Aphrodite and Louridas Panos, "Multilevel Petri nets for modeling and simulating organizational dynamic behavior", *Simulation & Gaming, Déc 1996, Vol.27, Issue 4.*
Tuomela, R. (1995). *The importance of us: A philosophy study of basic social notions.* Stanford, CA: Stanford University Press.
Telang, R., & Boatwright, P., & Mukhopadhyay, T. (2004). A mixture model for internet search-engine visits. *Journal of Marketing Research*, Vol. 61, No2, 206-214.
Upshaw, L. (1995). The keys to building cyberbrands. *Advertising Age,* Vol.66 Issue 22, p.18-22.
Viswanadham, N., & Srinivasa Raghavan, N.R. (2000). Performance analysis and design of supply chains: a Petri net approach, *Journal of Operational research Society*, *51*(10), 1158-1169.
Weigand, H., & De Moor, A., & Van den Heuvel, W.J. (2000). *Electronic Markets*. Vol. 10, Issue 4.
Weitz, W. (1998). *Workflow modelling for internet-based commerce: An approach based on high level Petri nets*. In G. Goos, J. Hartmanis, & J. van Leeuwen (Series Eds.) & W. Lamersdorf, & M. Merz (Vol.Eds.), *Lecture notes in Computer Science*, *Vol. 1402*. *Trends in Distributed Systems for Electronic Commerce* (pp.166-178). Heidelberg: Springer-Verlag.
Weyns Danny & Tom Holvoet (2002). "A Colored Petri Net for a Multi-Agent Application", Modeling Components, Objects and Agents, MOCA'02, Aarhus, Denmark August 26-27, in *Proceedings of MOCA'02, ISSN 0105-8517*, pp.121-140.

Wiederhold, G. (1992). Mediators in the architecture of future information systems. *IEEE Transactions on Computers*, 25(3), 38-49.

Wiemen, & S. Pingree (Eds.), Advancing communication science: Merging mass and interpersonal processes (pp. 110-134). Newbury Park, CA: Sage.

Williams, R.L., & Cothrel, J. (2000). Four Smart Ways To Run Online Communities. *Sloan Management Review*. Vol. 41, Issue 4.

Wooldridge, M. (2002). *An Introduction to MultiAgent Systems*. John Wileys & Sons. Chichester.

Wu, G. (1999). Perceived Interactivity and Attitude Toward Web Sites. In *Proceedings of the 1999 Conference of the American Academy of Advertising*, Marilyn S. Roberts, ed. Gainesville, FL: University of Florida.

Wu, D., & Ray, G;, & Geng, X., & Whinston, A. (2004). *Implications of Reduced Search Cost and Free Riding in E-Commerce*. Marketing Science, Vol.23, No 2, 255-262.

Zauberman, G. (2003). The intertemporal dynamics of consumer lock-in. *Journal of* Consumer Research, 30(3), 405-419.

ANNEXE 1 : Analyse comportementale des réseaux de Petri

L'évolution d'un réseau de Petri est représentée par l'évolution des marquages. Les notions de réseau de Petri vivant (toutes les transitions peuvent être franchies) et de réseau sans blocage sont importantes. Il y a blocage si aucune transition ne peut plus être franchie. Dans presque tous les cas, cela correspond à un système mal conçu ou mal modélisé. Nous allons donc détailler les propriétés usuelles des réseaux de Petri qui illustrent ces différents éléments.
Le vocabulaire utilisé dans les définitions ci-dessous regroupe les termes standards décrivant les réseaux de Petri utilisés dans la littérature (David et Alla, 1997 ; Diaz, 2001).

Propriétés usuelles :

1. Réseau de Petri borné :

Définition 1 : Une place Pi est dite **bornée** pour un marquage initial M_0 s'il existe un entier naturel k, tel que pour tout marquage accessible à partir de M_0, le nombre de jetons (ou marques) dans Pi est inférieur ou égal à k (on dit que Pi est k-bornée).
Définition 2 : Un réseau de Petri est **borné** pour un marquage initial M_0 si toutes les places sont bornées pour M_0 (le réseau est k-borné si toutes les places sont k-bornées).

Exemple de réseau de Petri non borné :

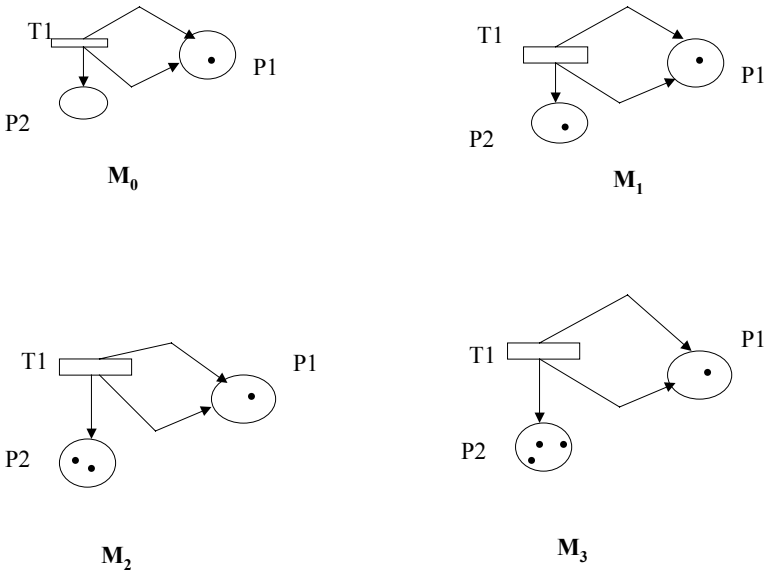


Figure 1: Réseau non borné

Ce réseau de Petri est non borné puisqu'à chaque transition, un jeton est ajouté dans la place P2, ce phénomène s'effectuant de manière illimitée.

2. Caractère vivant ou bloqué d'un réseau de Petri :

Le marquage d'un réseau de Petri évolue par franchissement de transitions. Lorsque certaines transitions ne sont plus franchissables, lorsque tout ou partie du réseau de Petri ne « fonctionne » plus, il y a probablement un problème dans la conception du système décrit. Les propriétés de vivacité et de blocage sont des notions relatives à ces problèmes.

Définition 3 : Une transition Tj est **vivante** pour un marquage initial M_0 si pour tout marquage accessible Mi $\in$ * M_0 (* M_0 étant l'ensemble des marquages accessibles à partir de M_0), il existe une séquence de franchissements S à partir de Mi qui contient la transition Tj. Autrement dit, quelle que soit l'évolution du système, il subsistera toujours une possibilité de franchir Tj.

Définition 4 : Un réseau de Petri est **vivant** pour un marquage initial M_0 si toutes ses transitions sont vivantes pour M_0.

Définition 5 : Une transition Tj est **quasi-vivante** pour un marquage initial M_0 , s'il existe au moins une séquence de franchissements qui contient la transition Tj à partir de M_0.
Autrement dit, une transition est quasi-vivante s'il existe une possibilité pour que cette transition soit franchie.

Définition 6 : Un réseau de Petri est **quasi-vivant** si toutes ses transitions sont quasi vivantes.

Définition 7 : Un **blocage** est un marquage tel qu'aucune transition n'est possible à partir de ce marquage.

Définition 8 : Un réseau de Petri est dit **sans blocage** pour un marquage initial M_0 si aucun marquage accessible Mi $\in$ * M_0 n'est un blocage.

Exemple :

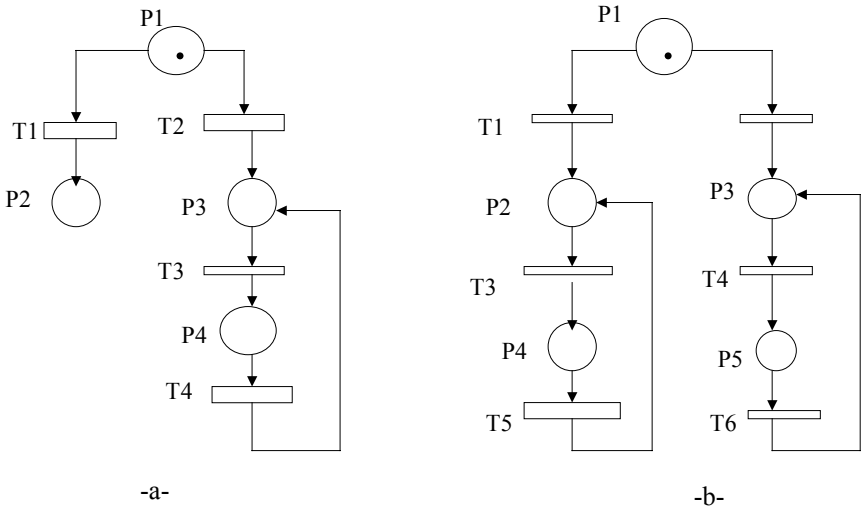


Figure 2: Réseaux de Pétri quasi-vivants
–a- avec blocage -b - sans blocage

Dans l'exemple de la figure 2 sont exposés deux réseaux de Petri quasi vivants. En effet, pour chaque transition, il existe une séquence de franchissement qui la contient, à partir du marquage initial indiqué. Cependant, dans le réseau –a- apparaît un conflit : si la première transition franchie est T1, alors les autres transitions ne seront plus jamais possibles, et si la première transition franchie est T2, alors T1 ne sera plus jamais possible.

Définition 9 : Un réseau de Petri a un **état d'accueil** M_a pour un marquage initial M_0 si pour tout marquage accessible Mi $\in$ * M_0 , il existe une séquence S_i telle que le franchissement de S_i à partir de M_i amène à M_a.

3. Conflit :

Définition 10 : Un **conflit structurel** correspond à un ensemble d'au moins deux transitions T1 et T2 qui ont une place d'entrée en commun. Il sera noté K .

Définition 11 : Un **conflit effectif** est l'existence d'un conflit structurel et d'un marquage M tel que le nombre de marques dans la place Pi est inférieur au nombre de transitions de sortie de Pi qui sont validées par M.

Exemple :

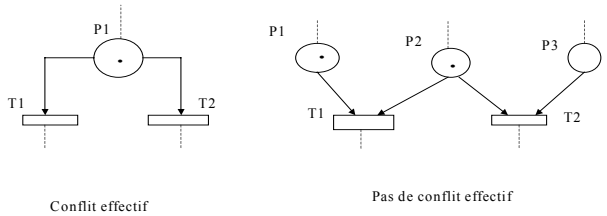


Figure 3: Illustration de la notion de conflit

L'exemple de la figure 3 montre que, dans le cas du conflit effectif, les deux transitions T1 et T2 sont possibles. La place P1 ne comportant qu'un seul jeton, il y a conflit entre les transitions T1 et T2 puisque le système ne définit pas de règle de choix entre les deux transitions possibles T1et T2. Dans le cas de droite, il n'y a pas de conflit effectif puisque le jeton de la place P1 peut franchir la transition T1 tandis que le jeton de la place P2 passera

soit par T1, soit par T2. Nous verrons ultérieurement que des règles peuvent être ajoutées aux arcs entre les places et les transitions pour solutionner ces problèmes de conflits.

4. Invariants :

A partir d'un marquage initial, le marquage d'un réseau de Petri peut évoluer par franchissements de transitions. S'il n'y a pas de blocage, le nombre de franchissements des transitions est illimité. Cependant, on ne pourra pas toujours atteindre n'importe quel marquage et on ne pourra pas toujours franchir n'importe quelle séquence de transitions. Des invariants permettent de caractériser certaines propriétés des marquages accessibles et des transitions franchissables, quelle que soit l'évolution du système.

Composante conservatrice :

Nous allons illustrer les propriétés usuelles des réseaux de Petri sur l'exemple suivant, qui comporte un jeton dans la place P1 lors du marquage initial :

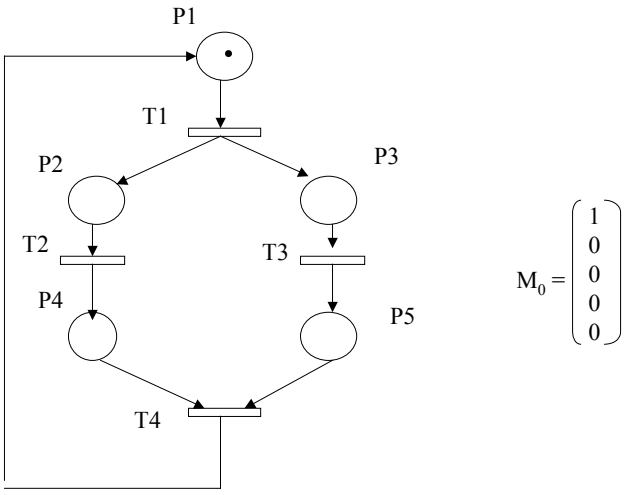
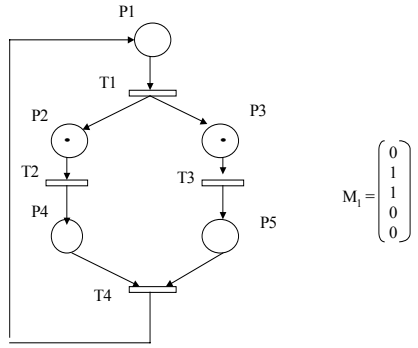
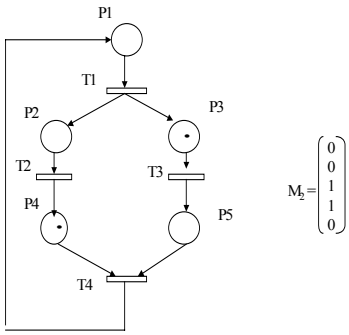


Figure 4: Réseau de Pétri avec son marquage initial

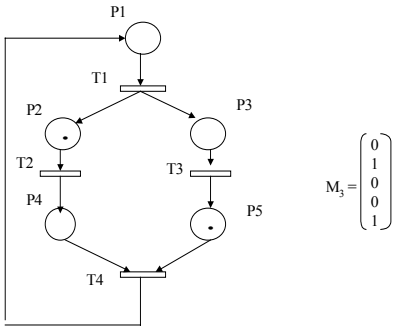
Les marquages accessibles de ce réseau sont les suivants :



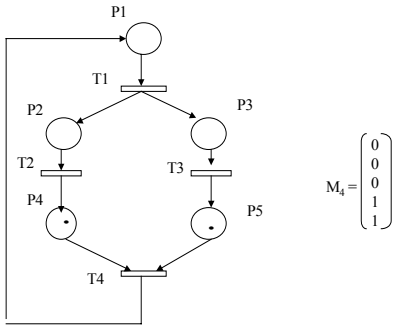
Marquage M1



Marquage M2



Marquage M3



Marquage M4

On constate dans ce réseau qu'il y a toujours une marque et une seule dans l'ensemble des places (P1, P2, P4). Si on désigne par m_i le nombre de marques (ou jetons) dans la place Pi, on peut alors écrire : $m_1 + m_2 + m_4 = 1$.

De façon similaire, on a $m_1 + m_3 + m_5 = 1$.

La somme des deux équations entraîne l'égalité suivante :

1 $m_1 + m_2 + m_3 + m_4 + m_5 = 2$

Ceci est un invariant qui est relatif à l'ensemble des places du réseau de Petri et qui signifie que la somme pondérée (poids 2 pour la place P1 et poids 1 pour les autres places) des marques dans le réseau de Petri est constante.

Définition 12 : Soit un réseau de Petri R, et P l'ensemble de ses places. On a un **invariant de marquage** (appelé aussi invariant linéaire de places) s'il existe un sous-ensemble de place $P' = \{P_1,P_2,\ldots,P_r\}$ inclus dans P et un vecteur de pondération $(q_1,q_2,\ldots,q_r)$ dont tous les poids q_i sont des nombres entiers positifs, tels que :

$q_1 M(P_1) + q_2 M(P_2) + \ldots + q_r M(P_r)$ = constante, pour tout marquage $M \in * M_0$
avec $M(P_r) = m_r$

L'ensemble de places P' est une composante **conservatrice.**
En règle générale, une composante conservatrice a une signification physique. Par exemple, un système est dans un et un seul état à la fois, ou un nombre d'entités (jetons) se conserve.

Composante répétitive:
Dans l'exemple de réseau de Petri de la figure 8, les séquences de franchissements qui sont possibles à partir du marquage initial M_0 sont les suivantes : T1, T1T2, T1T2T3, T1T2T3T4, T1T2T3T4T1, etc…La séquence de franchissement T1T2T3T4 est particulière parce qu'elle ramène à l'état initial M_0. On pourra donc répéter ces séquences. C'est une séquence **répétitive.**

Définition 13 : Soit un réseau de Petri R et T l'ensemble de ses transitions. Soit S une séquence de franchissements répétitive telle que les transitions qui apparaissent dans S sont définies par le sous-ensemble T' = {T1, T2, …, Tr} inclus dans T. L'ensemble des transitions T' est une **composante répétitive**.
Le réseau R est dit **répétitif** si et seulement si T est une composante répétitive.

Pour le réseau de Petri ci-dessus, {T1, T2, T3, T4} est une composante répétitive. Cette composante comprenant toutes les transitions, le réseau de Petri est répétitif.
Les séquences répétitives donnent une idée du comportement « cyclique » d'un réseau.

Théorie générale des invariants :

Considérons un vecteur de pondération des places
F = (q1, q2, …, qn) qi nombre entier ≥ 0
qi est le poids associé à la place Pi
P(F) : ensemble des places dont le poids est non nul
P(F) est un sous ensemble de P

B est une composante conservative si et seulement si il existe un vecteur de pondération F tel que :

P(F) = B et F.W = 0

(En effet on a : Mi = Mo + W.$\underline{S}$ pour tout marquage Mi atteignable par une séquence de franchissements S

On en déduit :

F.Mi = F.Mo + F.W.$\underline{S}$

Si F.W= 0, on a : F.Mi = F.Mo pour tout S)

F.Mi est un **invariant de marquage** : le nombre de marques dans l'ensemble de places P(F), pondéré par le vecteur F, est invariant et constant.

Un vecteur F ≥0 qui est solution de F.W = 0 est appelé **P-semi-flot**

Caractérisation des propriétés d'un réseau de Petri à l'aide d'un graphe d'accessibilité fini :

L'un des intérêts majeurs des modèles formels des réseaux de Petri est la possibilité de définir le comportement d'un système et de développer des algorithmes de vérification de ses propriétés. La vérification des propriétés du réseau s'effectue grâce au graphe des marquages du réseau, ou graphe d'états.

Le graphe d'états d'un réseau de Petri représente tous les états accessibles du système et toutes les actions possibles à partir de ces états. Les méthodes d'analyse basées sur ce graphe sont appelées classiquement méthodes comportementales.

Le graphe G s'obtient de la manière suivante :

- pour chaque M_i à partir de M_0, trouver les transitions T_i franchissables à partir de M_i ;
- sélectionner et tirer chaque T_i (à partir du même marquage) ; trouver le marquage suivant M_i' ; construire le nouveau sommet si M_i' est différent des marquages précédents ; noter l'arc correspondant ;
- tant qu'il existe des marquages non considérés, continuer.

Exemple :

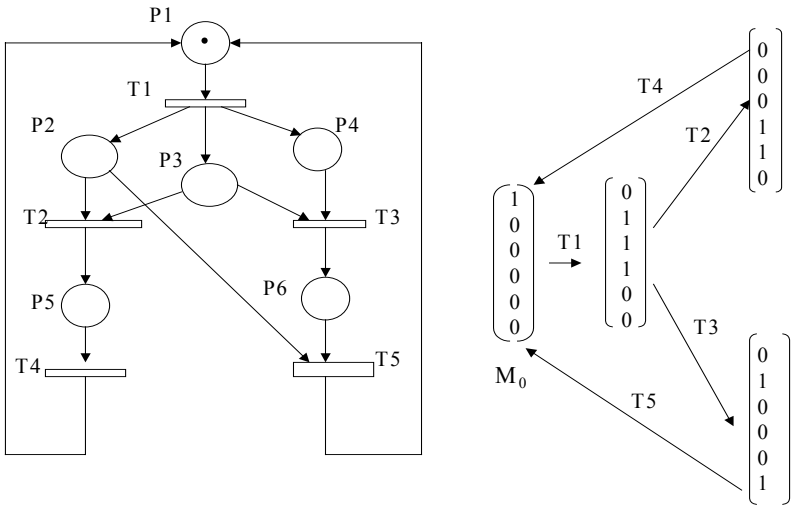


Figure 5: Réseau de Pétri et son graphe de marquage

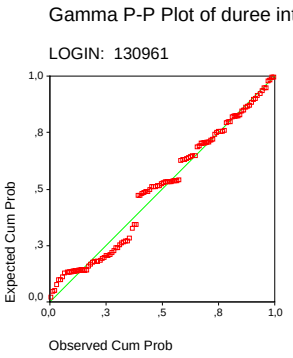
Ce graphe de marquage permet de trouver toutes les propriétés de ce réseau de Petri. On voit par exemple qu'il est vivant, réinitialisable et que T1T2T4 et T1T3T5 sont des séquences répétitives.

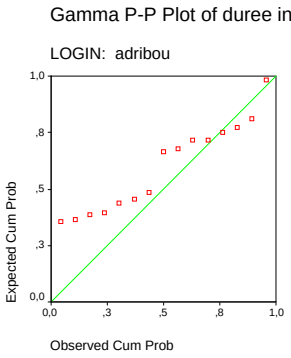
Tous les points développés dans l'annexe 1 sont principalement utilisés dans le cadre de constructions de réseaux , de design de systèmes ou de gestion de production (les utilisations les plus fréquentes concernent l'informatique ou la recherche opérationnelle). Nous les avons présenté ici afin d'avoir une vue complète sur les réseaux de Petri mais ils ne seront pas intégrés dans les applications développées dans la thèse, où nous utiliserons les réseaux de Petri pour modéliser des processus et en déterminer les performances.

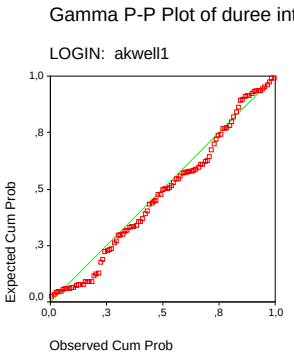
ANNEXE 2

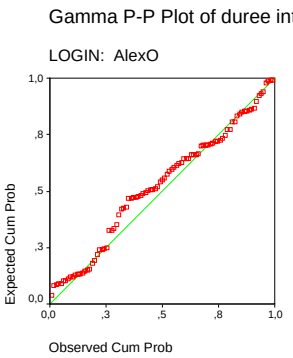
PP plots des lois Gamma, Log-normale et Weibull appliquées à la variable durée interconnexion de chacun des 135 agents A :

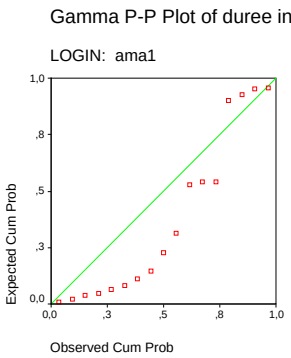
PP PLOT LOI GAMMA :

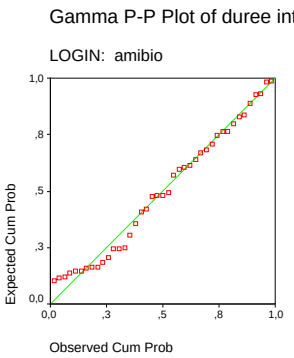


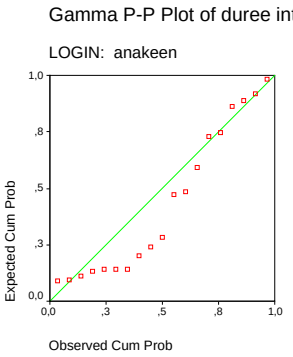


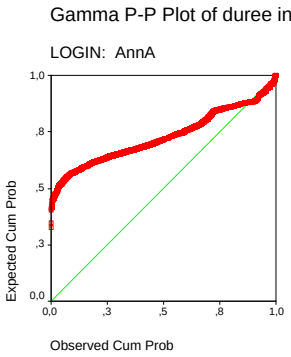


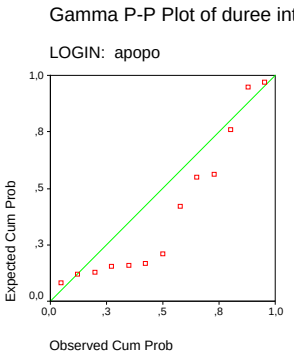






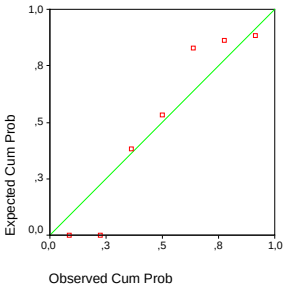






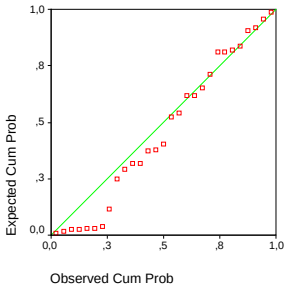
Gamma P-P Plot of duree interc

LOGIN: ashley85



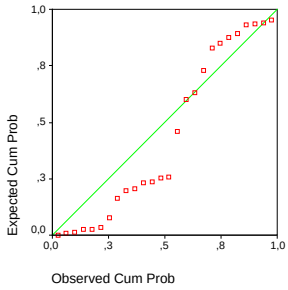
Gamma P-P Plot of duree interc

LOGIN: Aude



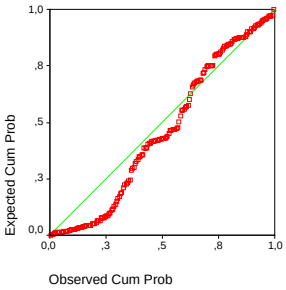
Gamma P-P Plot of duree interc

LOGIN: badir



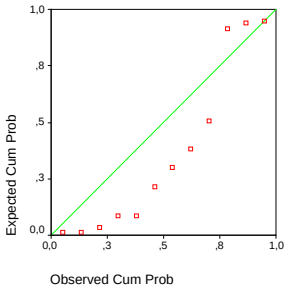
Gamma P-P Plot of duree interc

LOGIN: bgougay



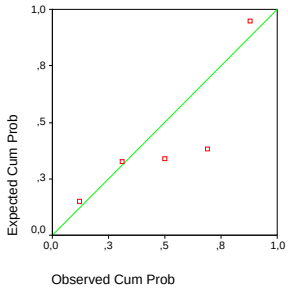
Gamma P-P Plot of duree interc

LOGIN: blondi



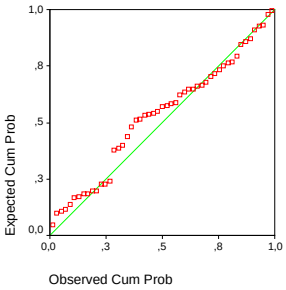
Gamma P-P Plot of duree interc

LOGIN: bontze



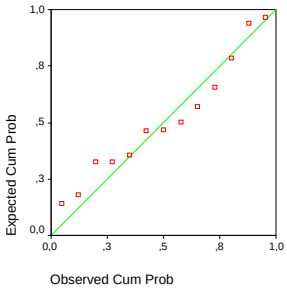
Gamma P-P Plot of duree interc

LOGIN: bpdl



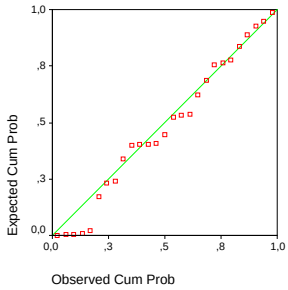
Gamma P-P Plot of duree interc

LOGIN: brusson1



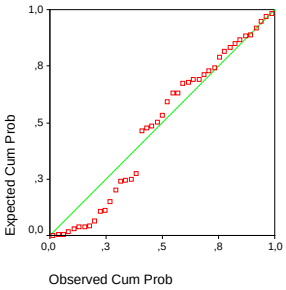
Gamma P-P Plot of duree interc

LOGIN: bruyere



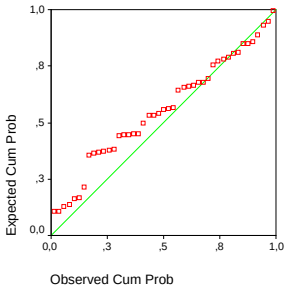
Gamma P-P Plot of duree interc

LOGIN: bschunke



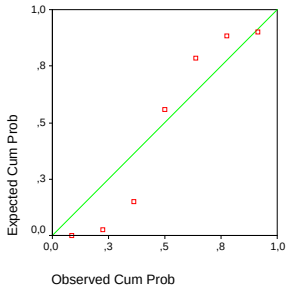
Gamma P-P Plot of duree interc

LOGIN: carolala



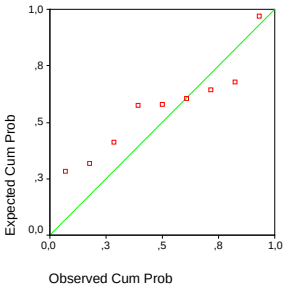
Gamma P-P Plot of duree interc

LOGIN: cathou1



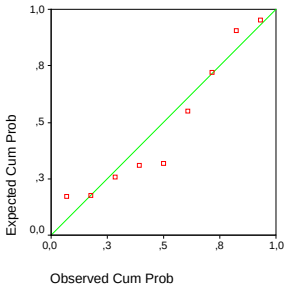
Gamma P-P Plot of duree interc

LOGIN: CATHOU25



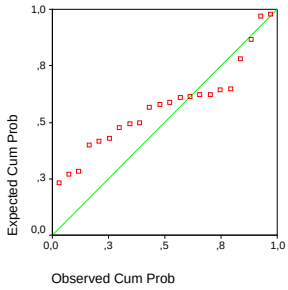
Gamma P-P Plot of duree interc

LOGIN: cclerget



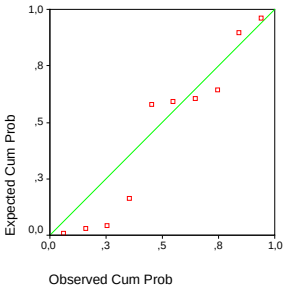
Gamma P-P Plot of duree interc

LOGIN: Céline21



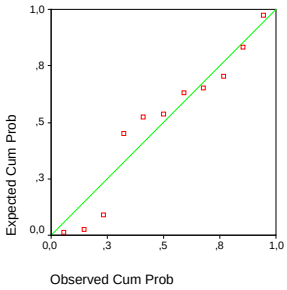
Gamma P-P Plot of duree interc

LOGIN: chap's



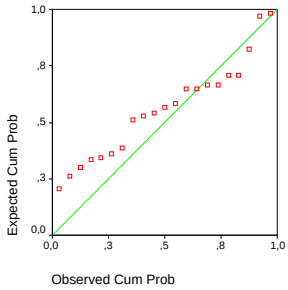
Gamma P-P Plot of duree interc

LOGIN: Chrispi1



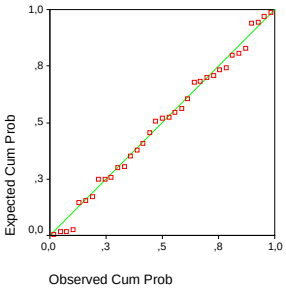
Gamma P-P Plot of duree interc

LOGIN: clemisa



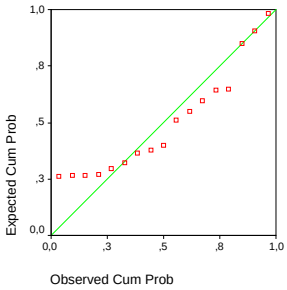
Gamma P-P Plot of duree interc

LOGIN: Cypher



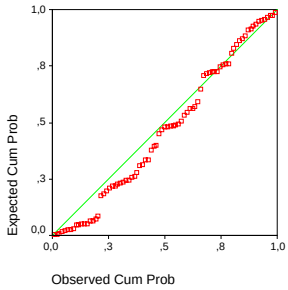
Gamma P-P Plot of duree interc

LOGIN: DALMATE



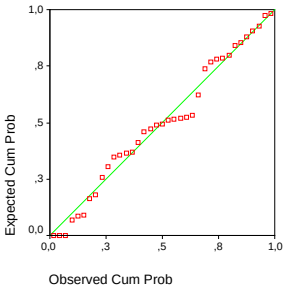
Gamma P-P Plot of duree interc

LOGIN: Delatte



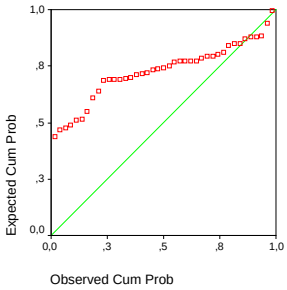
Gamma P-P Plot of duree interc

LOGIN: delillle



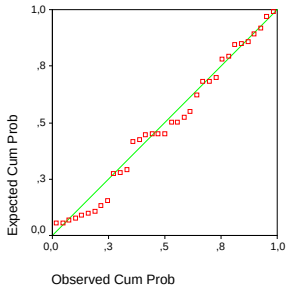
Gamma P-P Plot of duree interc

LOGIN: didine77



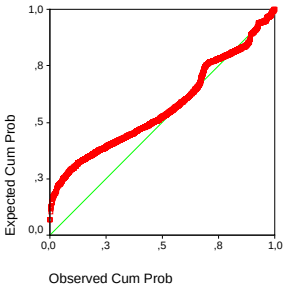
Gamma P-P Plot of duree interc

LOGIN: Dinou



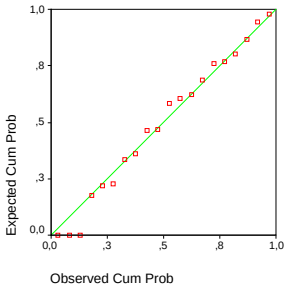
Gamma P-P Plot of duree interc

LOGIN: djoule



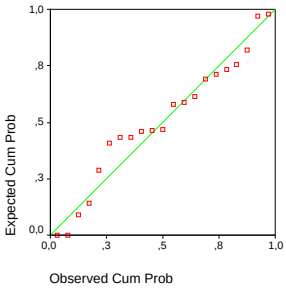
Gamma P-P Plot of duree interc

LOGIN: dolores3



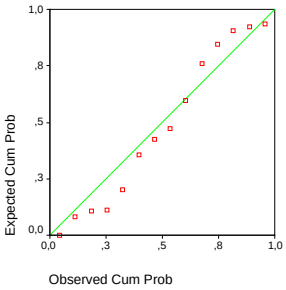
Gamma P-P Plot of duree interc

LOGIN: doppel



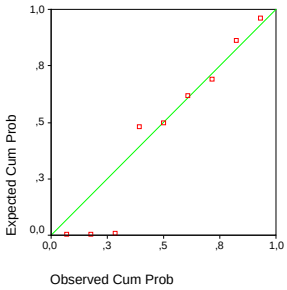
Gamma P-P Plot of duree interc

LOGIN: doppel2



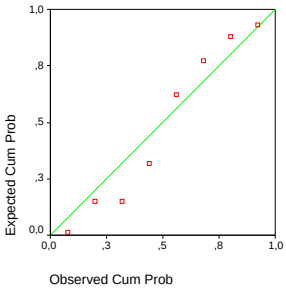
Gamma P-P Plot of duree interc

LOGIN: dperrot



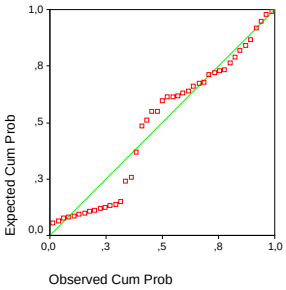
Gamma P-P Plot of duree interc

LOGIN: dzhouari



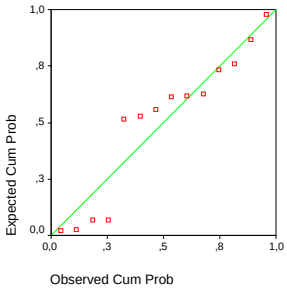
Gamma P-P Plot of duree interc

LOGIN: eb95



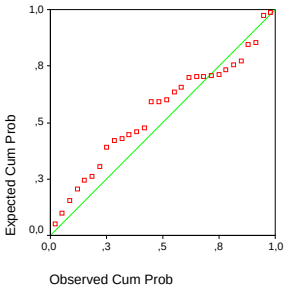
Gamma P-P Plot of duree interc

LOGIN: eb96



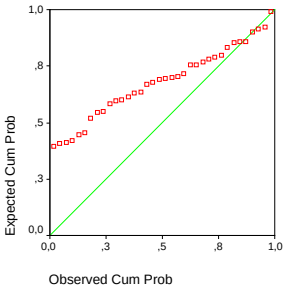
Gamma P-P Plot of duree interc

LOGIN: ehran



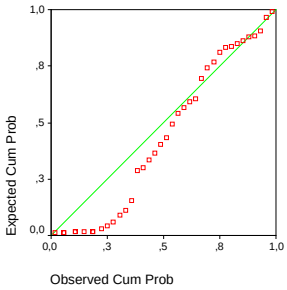
Gamma P-P Plot of duree interc

LOGIN: Emilie



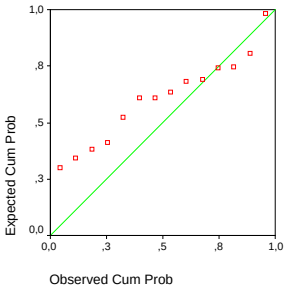
Gamma P-P Plot of duree interc

LOGIN: estia



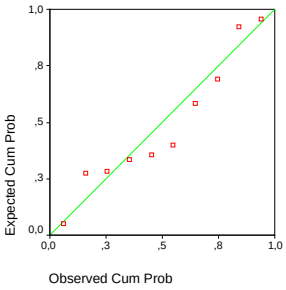
Gamma P-P Plot of duree interc

LOGIN: eusebio1



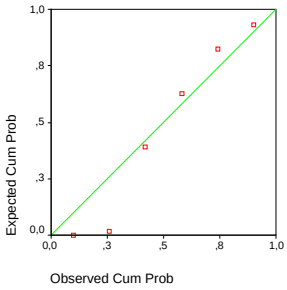
Gamma P-P Plot of duree interc

LOGIN: Fab007



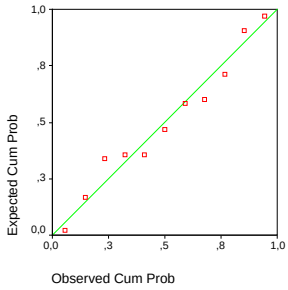
Gamma P-P Plot of duree interc

LOGIN: fabby93



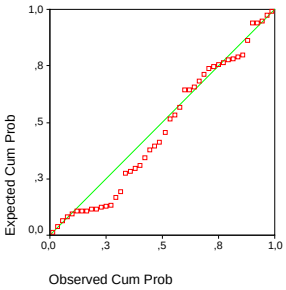
Gamma P-P Plot of duree interc

LOGIN: fabou



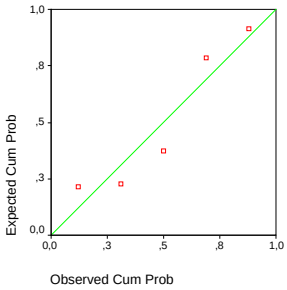
Gamma P-P Plot of duree interc

LOGIN: fallouis



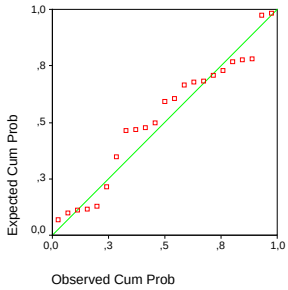
Gamma P-P Plot of duree interc

LOGIN: fdrugeat



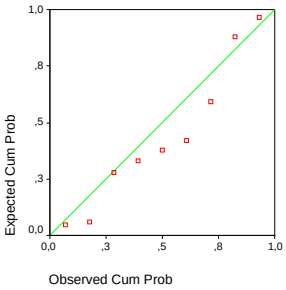
Gamma P-P Plot of duree interc

LOGIN: fevrard



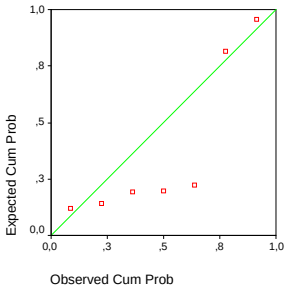
Gamma P-P Plot of duree interc

LOGIN: FOURNIAL



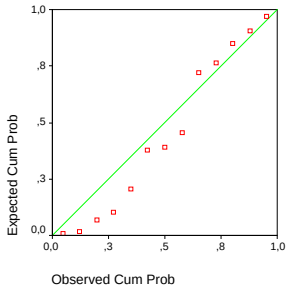
Gamma P-P Plot of duree interc

LOGIN: Fruor



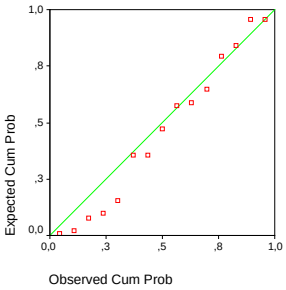
Gamma P-P Plot of duree interc

LOGIN: frzpower



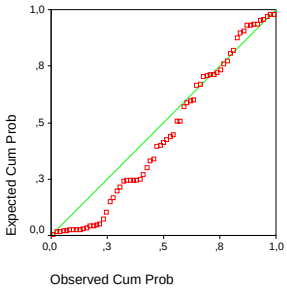
Gamma P-P Plot of duree interc

LOGIN: GaëlleM



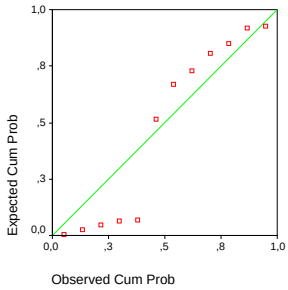
Gamma P-P Plot of duree interc

LOGIN: gallouin



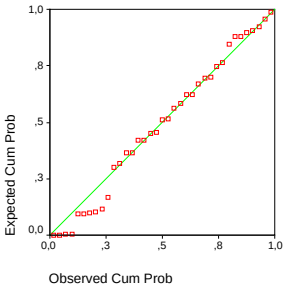
Gamma P-P Plot of duree interc

LOGIN: Ginie31



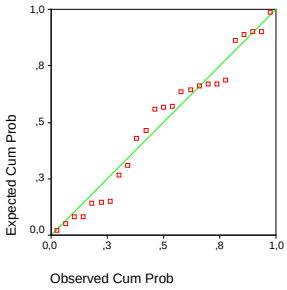
Gamma P-P Plot of duree interc

LOGIN: gougay



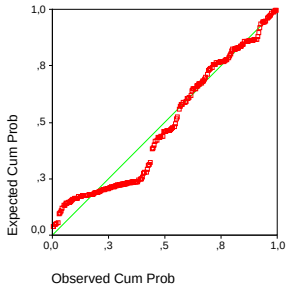
Gamma P-P Plot of duree interc

LOGIN: Greensky



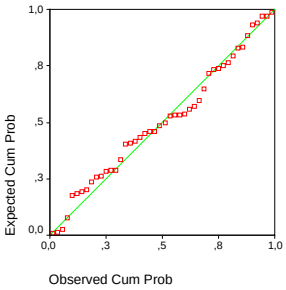
Gamma P-P Plot of duree interc

LOGIN: guix



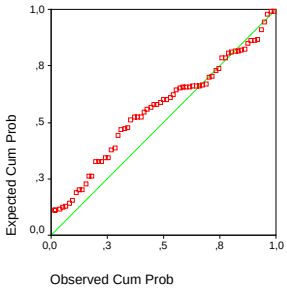
Gamma P-P Plot of duree interc

LOGIN: gzing



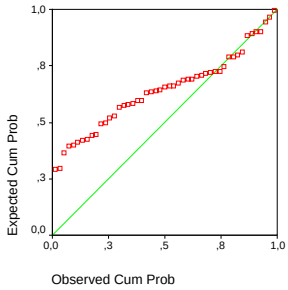
Gamma P-P Plot of duree interc

LOGIN: hedi.dc



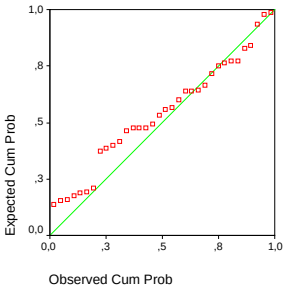
Gamma P-P Plot of duree interc

LOGIN: HERPIN



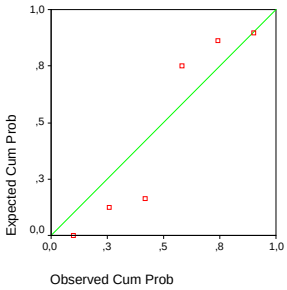
Gamma P-P Plot of duree interc

LOGIN: hutch



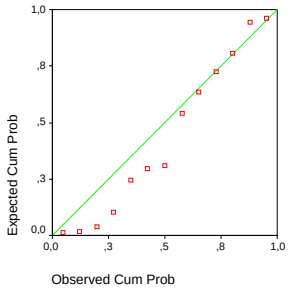
Gamma P-P Plot of duree interc

LOGIN: j-kolb



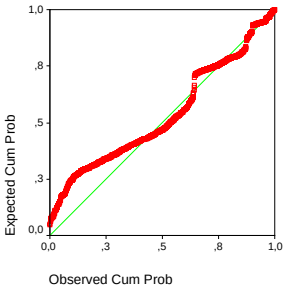
Gamma P-P Plot of duree interc

LOGIN: jeanbask



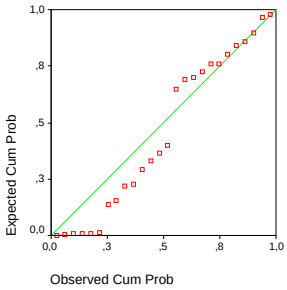
Gamma P-P Plot of duree interc

LOGIN: jeannot3



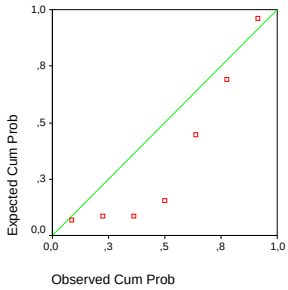
Gamma P-P Plot of duree interc

LOGIN: jeda



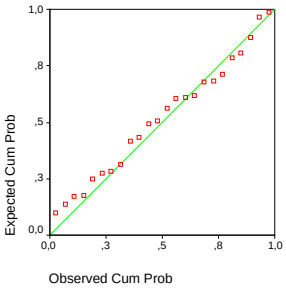
Gamma P-P Plot of duree interc

LOGIN: JIMASA



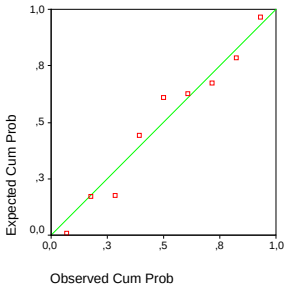
Gamma P-P Plot of duree interc

LOGIN: joewj



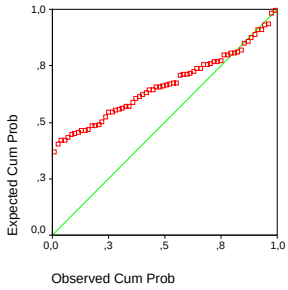
Gamma P-P Plot of duree interc

LOGIN: josetta



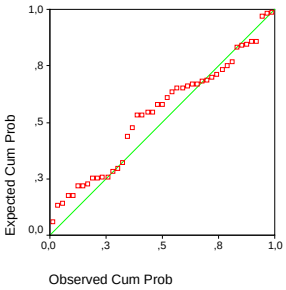
Gamma P-P Plot of duree interc

LOGIN: juboubou



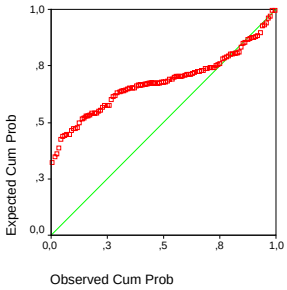
Gamma P-P Plot of duree interc

LOGIN: juju75



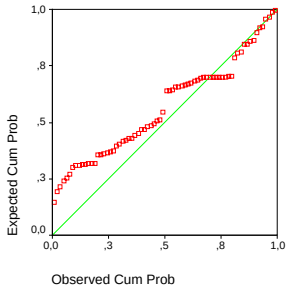
Gamma P-P Plot of duree interc

LOGIN: kabotine



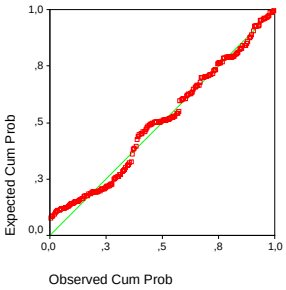
Gamma P-P Plot of duree interc

LOGIN: Kathrin



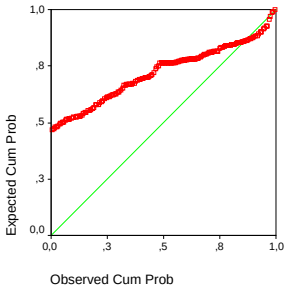
Gamma P-P Plot of duree interc

LOGIN: Kifran



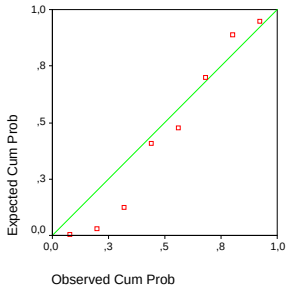
Gamma P-P Plot of duree interc

LOGIN: lbarbier



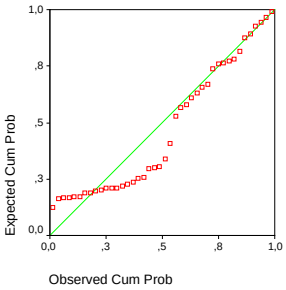
Gamma P-P Plot of duree interc

LOGIN: leleu5



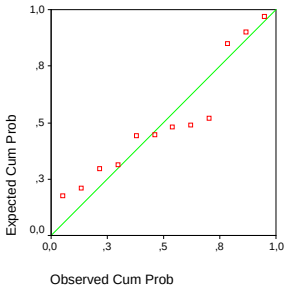
Gamma P-P Plot of duree interc

LOGIN: levieux



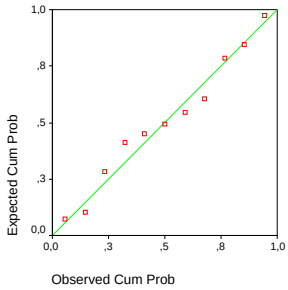
Gamma P-P Plot of duree interc

LOGIN: lilimimi



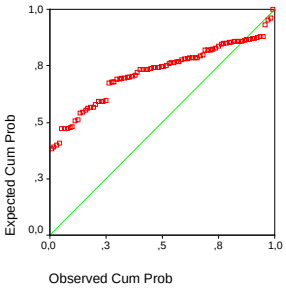
Gamma P-P Plot of duree interc

LOGIN: LoïcT



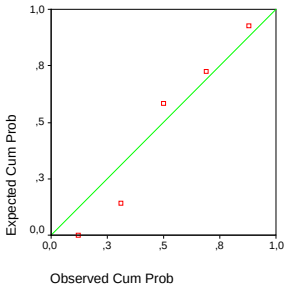
Gamma P-P Plot of duree interc

LOGIN: Loquet



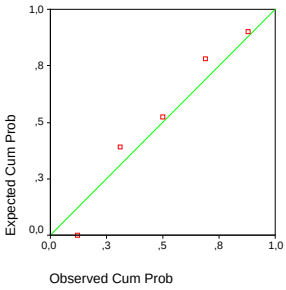
Gamma P-P Plot of duree interc

LOGIN: luberdov



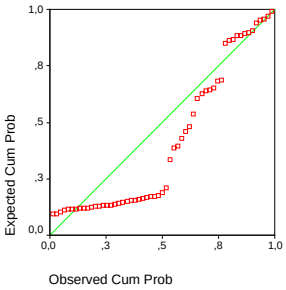
Gamma P-P Plot of duree interc

LOGIN: ludovicb



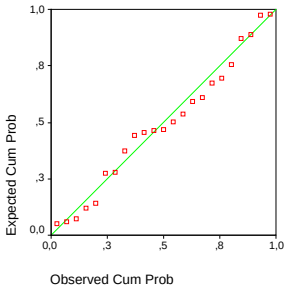
Gamma P-P Plot of duree interc

LOGIN: m@hy



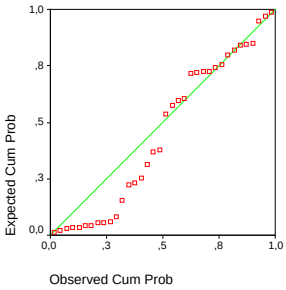
Gamma P-P Plot of duree interc

LOGIN: manwe



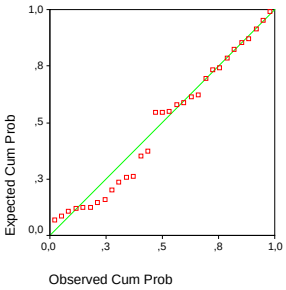
Gamma P-P Plot of duree interc

LOGIN: mary_ce



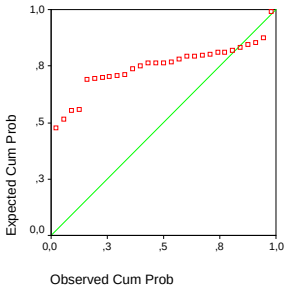
Gamma P-P Plot of duree interc

LOGIN: Mayfair



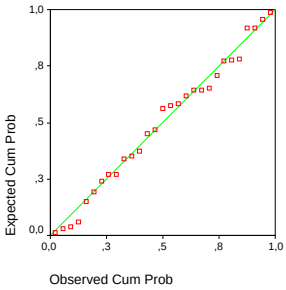
Gamma P-P Plot of duree interc

LOGIN: mekiela



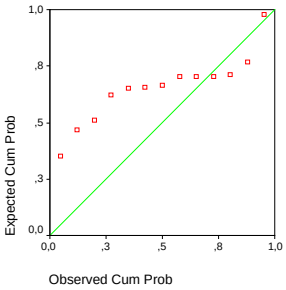
Gamma P-P Plot of duree interc

LOGIN: Michel



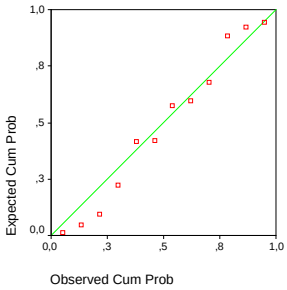
Gamma P-P Plot of duree interc

LOGIN: micka74



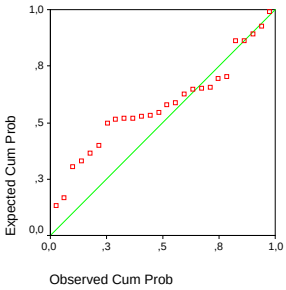
Gamma P-P Plot of duree interc

LOGIN: mike_s



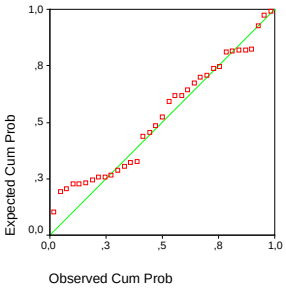
Gamma P-P Plot of duree interc

LOGIN: miloumoi



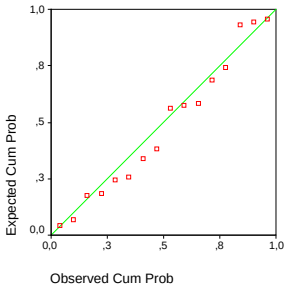
Gamma P-P Plot of duree interc

LOGIN: minious



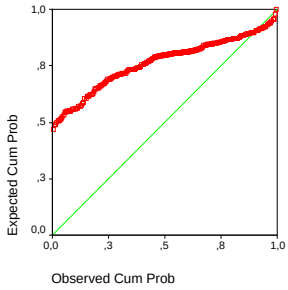
Gamma P-P Plot of duree interc

LOGIN: modjo-13



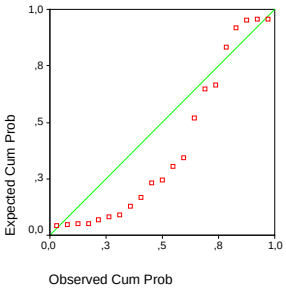
Gamma P-P Plot of duree interc

LOGIN: mperrot



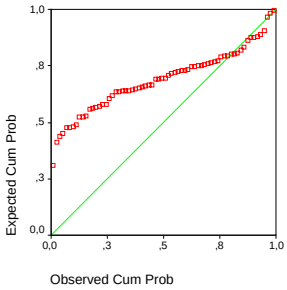
Gamma P-P Plot of duree interc

LOGIN: Munier



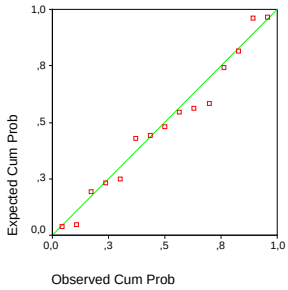
Gamma P-P Plot of duree interc

LOGIN: Naemesis



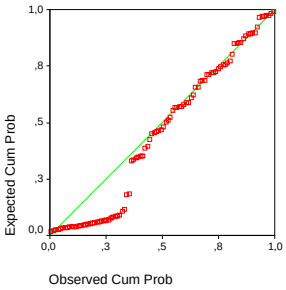
Gamma P-P Plot of duree interc

LOGIN: Nathalie B



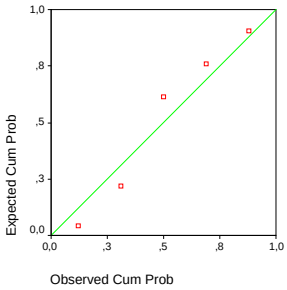
Gamma P-P Plot of duree interc

LOGIN: nhai07



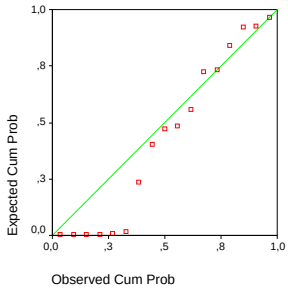
Gamma P-P Plot of duree interc

LOGIN: nickus



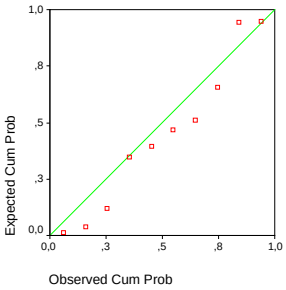
Gamma P-P Plot of duree interc

LOGIN: ninie12



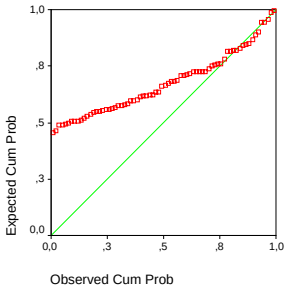
Gamma P-P Plot of duree interc

LOGIN: noel_net



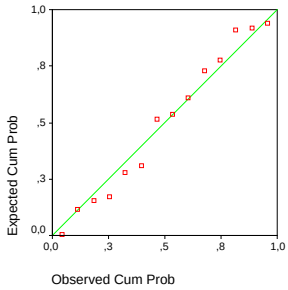
Gamma P-P Plot of duree interc

LOGIN: philia



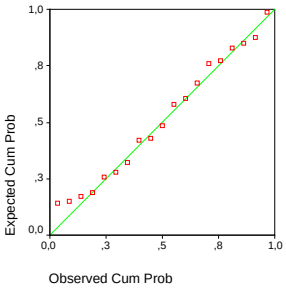
Gamma P-P Plot of duree interc

LOGIN: picardjf



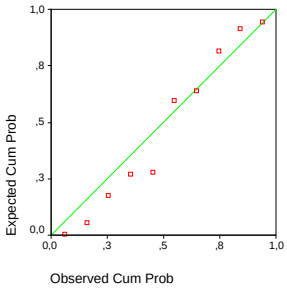
Gamma P-P Plot of duree interc

LOGIN: popcorn



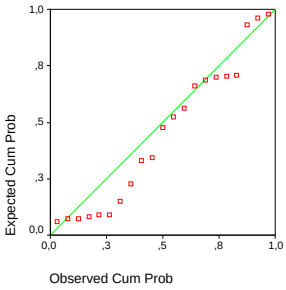
Gamma P-P Plot of duree interc

LOGIN: poupide



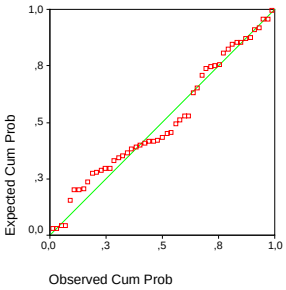
Gamma P-P Plot of duree interc

LOGIN: prioton



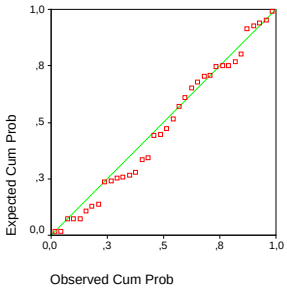
Gamma P-P Plot of duree interc

LOGIN: prozaiik



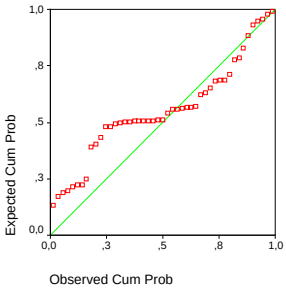
Gamma P-P Plot of duree interc

LOGIN: quentin



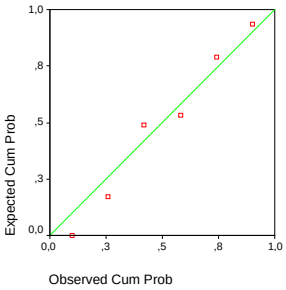
Gamma P-P Plot of duree interc

LOGIN: RAPH



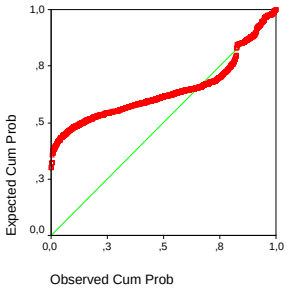
Gamma P-P Plot of duree interc

LOGIN: relindis



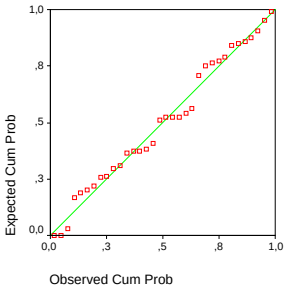
Gamma P-P Plot of duree interc

LOGIN: rlaval



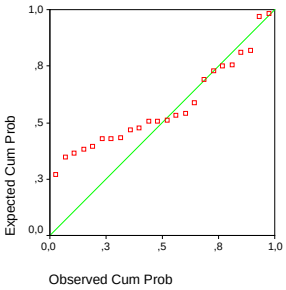
Gamma P-P Plot of duree interc

LOGIN: romain37



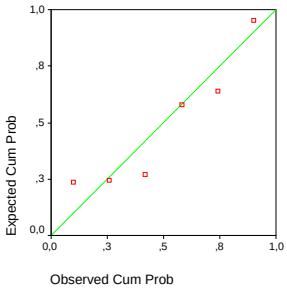
Gamma P-P Plot of duree interc

LOGIN: sandor



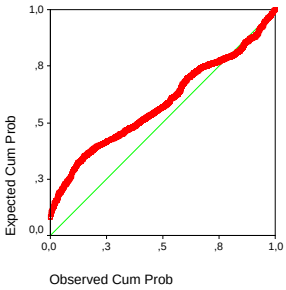
Gamma P-P Plot of duree interc

LOGIN: sebulba9



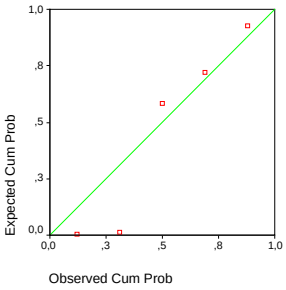
Gamma P-P Plot of duree interc

LOGIN: semery



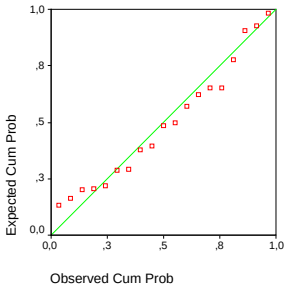
Gamma P-P Plot of duree interc

LOGIN: sicart



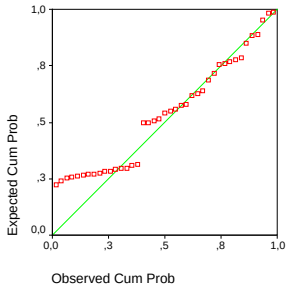
Gamma P-P Plot of duree interc

LOGIN: sophynet



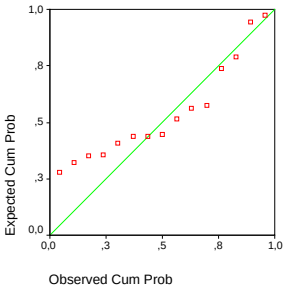
Gamma P-P Plot of duree interc

LOGIN: Souris



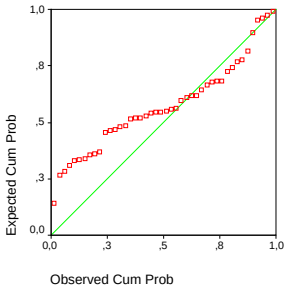
Gamma P-P Plot of duree interc

LOGIN: starchen



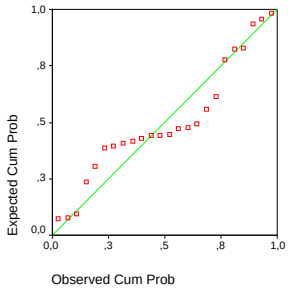
Gamma P-P Plot of duree interc

LOGIN: Stèph



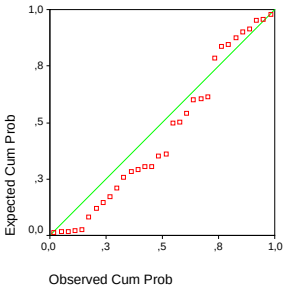
Gamma P-P Plot of duree interc

LOGIN: stumpf



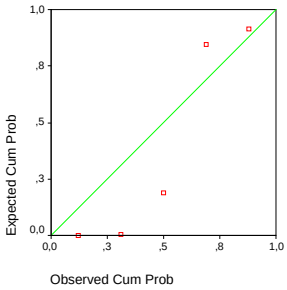
Gamma P-P Plot of duree interc

LOGIN: Sylvie4



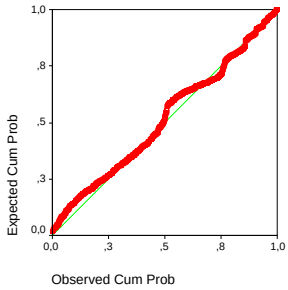
Gamma P-P Plot of duree interc

LOGIN: Tantan



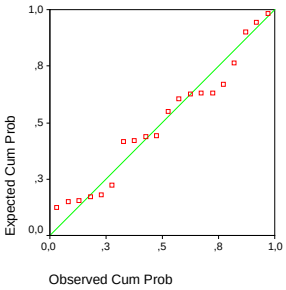
Gamma P-P Plot of duree interc

LOGIN: Thibaut



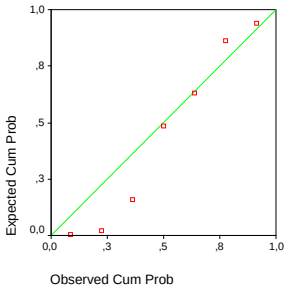
Gamma P-P Plot of duree interc

LOGIN: Tonio2k



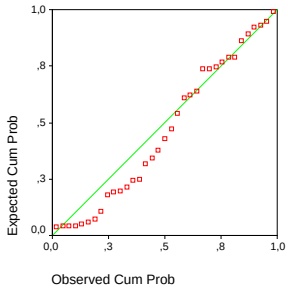
Gamma P-P Plot of duree interc

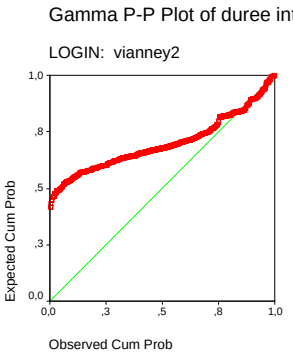
LOGIN: toto2

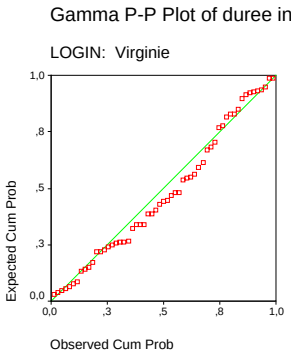


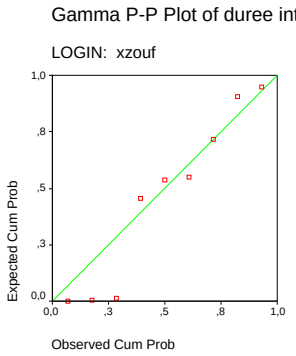
Gamma P-P Plot of duree interc

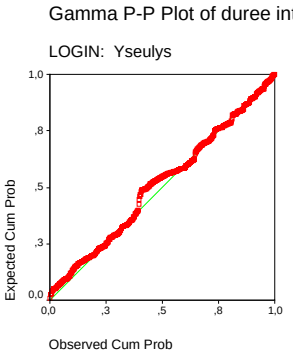
LOGIN: tranb

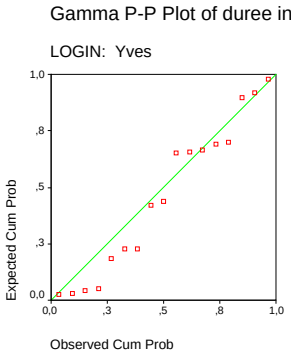


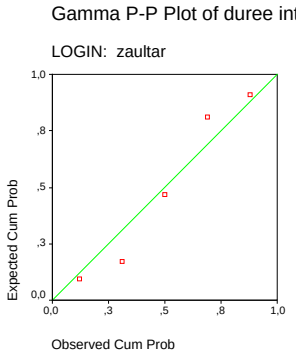




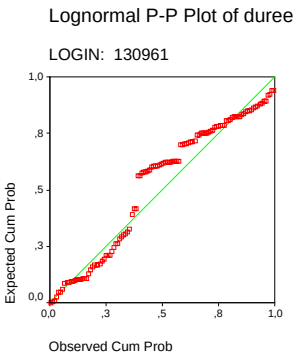


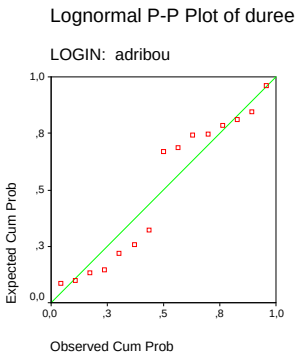


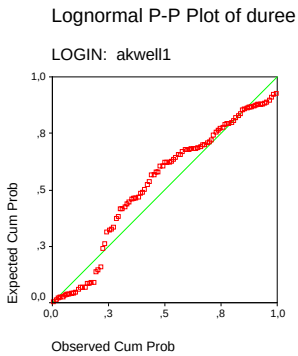




PP PLOT LOI LOG-NORMALE :

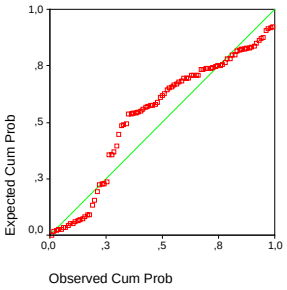






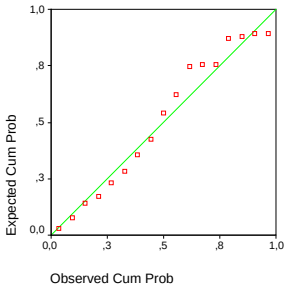
Lognormal P-P Plot of duree inte

LOGIN: AlexO



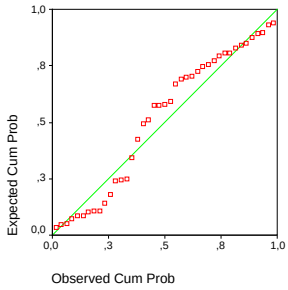
Lognormal P-P Plot of duree inte

LOGIN: ama1



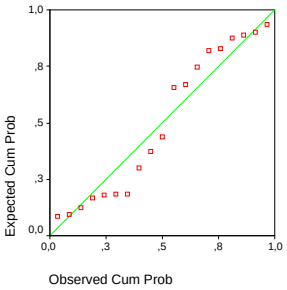
Lognormal P-P Plot of duree inte

LOGIN: amibio



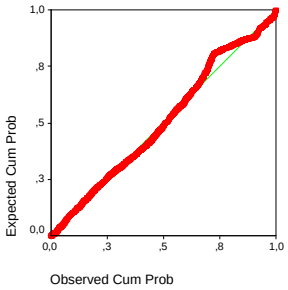
Lognormal P-P Plot of duree inte

LOGIN: anakeen



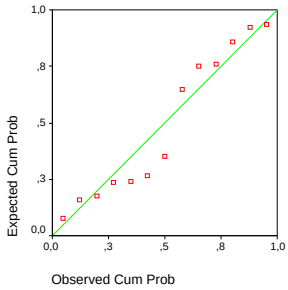
Lognormal P-P Plot of duree inte

LOGIN: AnnA



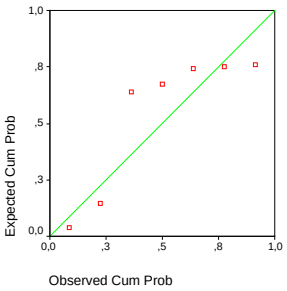
Lognormal P-P Plot of duree inte

LOGIN: apopo



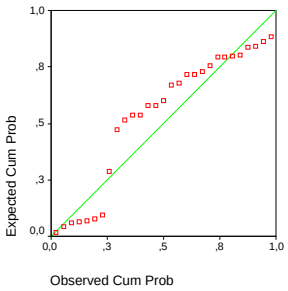
Lognormal P-P Plot of duree inte

LOGIN: ashley85



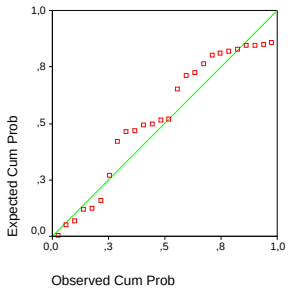
Lognormal P-P Plot of duree inte

LOGIN: Aude



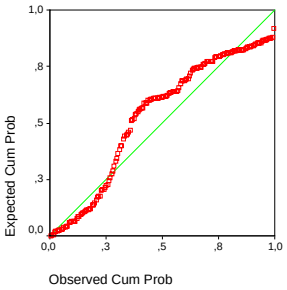
Lognormal P-P Plot of duree inte

LOGIN: badir



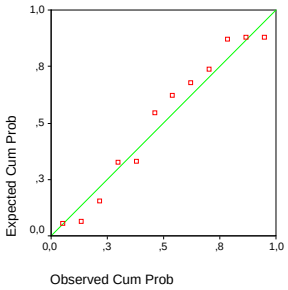
Lognormal P-P Plot of duree inte

LOGIN: bgougay



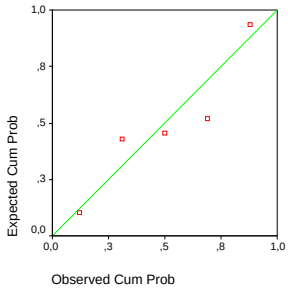
Lognormal P-P Plot of duree inte

LOGIN: blondi



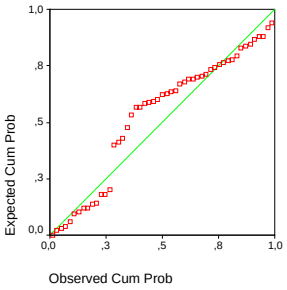
Lognormal P-P Plot of duree inte

LOGIN: bontze



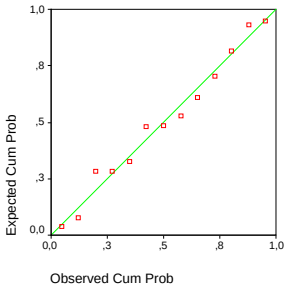
Lognormal P-P Plot of duree inte

LOGIN: bpdl



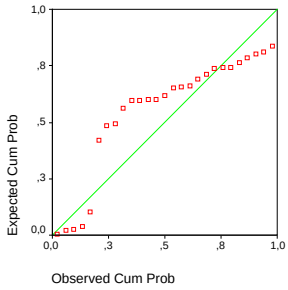
Lognormal P-P Plot of duree inte

LOGIN: brusson1



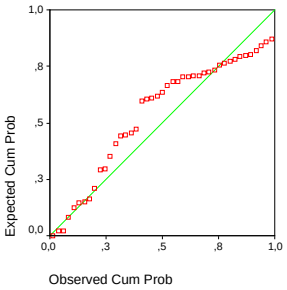
Lognormal P-P Plot of duree inte

LOGIN: bruyere



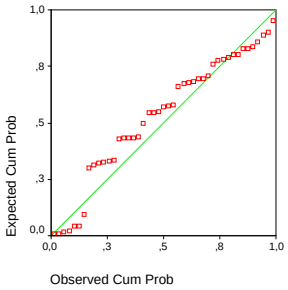
Lognormal P-P Plot of duree inte

LOGIN: bschunke



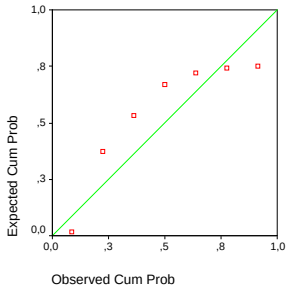
Lognormal P-P Plot of duree inte

LOGIN: carolala



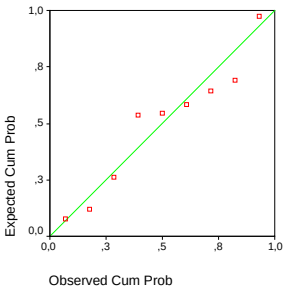
Lognormal P-P Plot of duree inte

LOGIN: cathou1



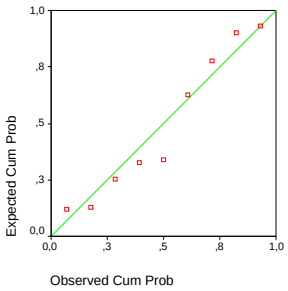
Lognormal P-P Plot of duree inte

LOGIN: CATHOU25



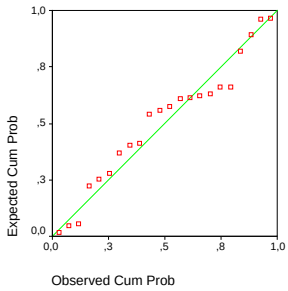
Lognormal P-P Plot of duree inte

LOGIN: cclerget



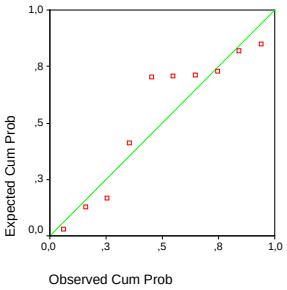
Lognormal P-P Plot of duree inte

LOGIN: Céline21



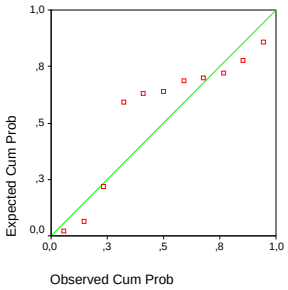
Lognormal P-P Plot of duree inte

LOGIN: chap's



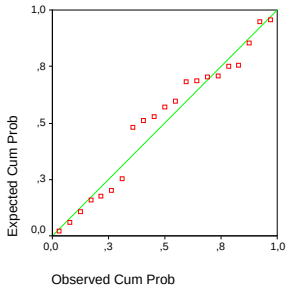
Lognormal P-P Plot of duree inte

LOGIN: Chrispi1



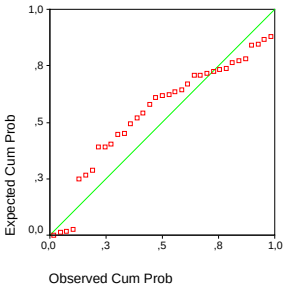
Lognormal P-P Plot of duree inte

LOGIN: clemisa



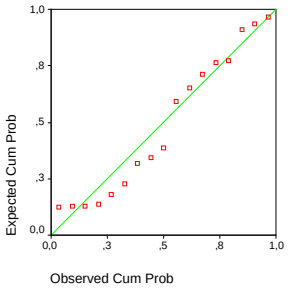
Lognormal P-P Plot of duree inte

LOGIN: Cypher



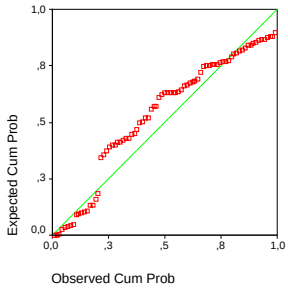
Lognormal P-P Plot of duree inte

LOGIN: DALMATE



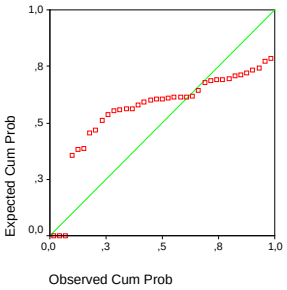
Lognormal P-P Plot of duree inte

LOGIN: Delatte



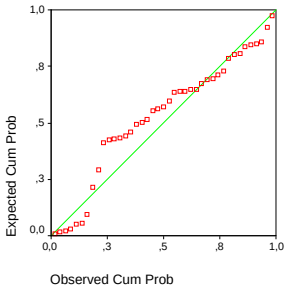
Lognormal P-P Plot of duree inte

LOGIN: delillle



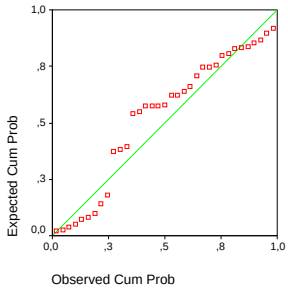
Lognormal P-P Plot of duree inte

LOGIN: didine77



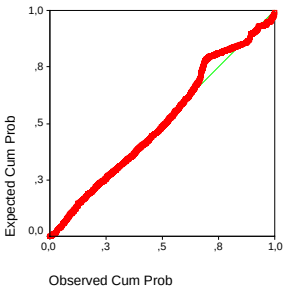
Lognormal P-P Plot of duree inte

LOGIN: Dinou



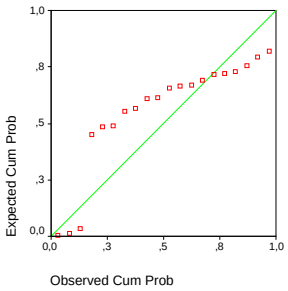
Lognormal P-P Plot of duree inte

LOGIN: djoule



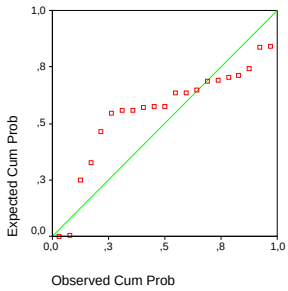
Lognormal P-P Plot of duree inte

LOGIN: dolores3



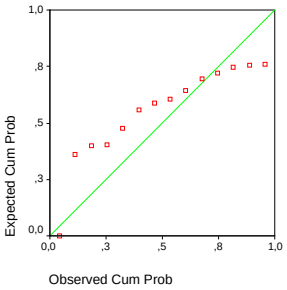
Lognormal P-P Plot of duree inte

LOGIN: doppel



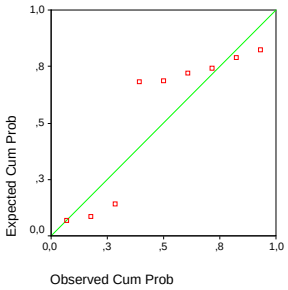
Lognormal P-P Plot of duree inte

LOGIN: doppel2



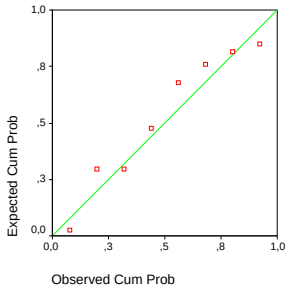
Lognormal P-P Plot of duree inte

LOGIN: dperrot



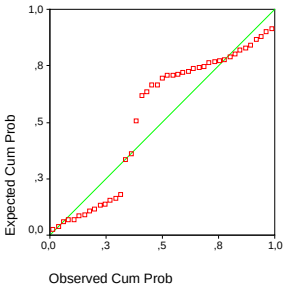
Lognormal P-P Plot of duree inte

LOGIN: dzhouari



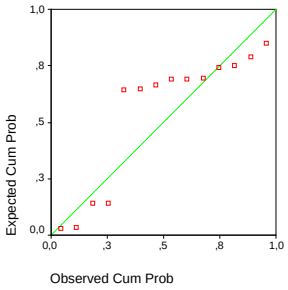
Lognormal P-P Plot of duree inte

LOGIN: eb95



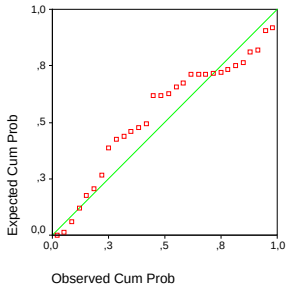
Lognormal P-P Plot of duree inte

LOGIN: eb96



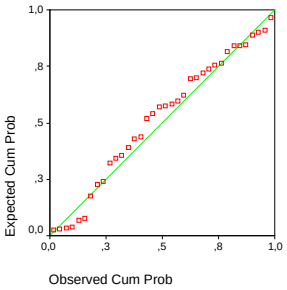
Lognormal P-P Plot of duree inte

LOGIN: ehran



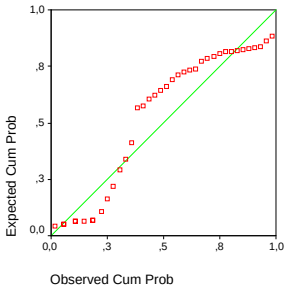
Lognormal P-P Plot of duree inte

LOGIN: Emilie



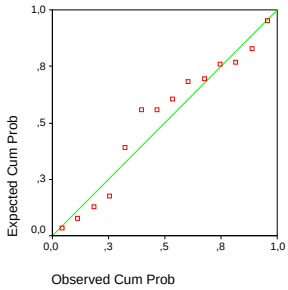
Lognormal P-P Plot of duree inte

LOGIN: estia



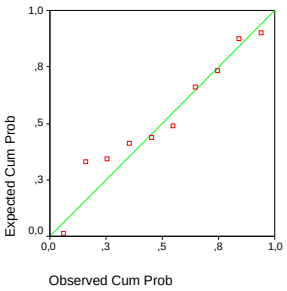
Lognormal P-P Plot of duree inte

LOGIN: eusebio1



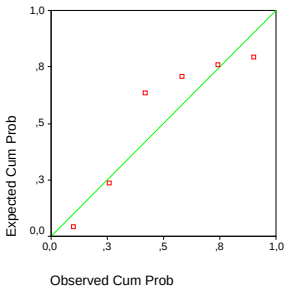
Lognormal P-P Plot of duree inte

LOGIN: Fab007



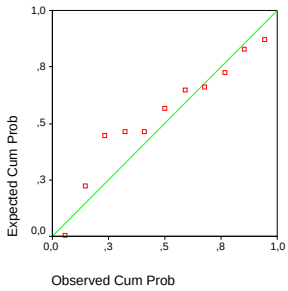
Lognormal P-P Plot of duree inte

LOGIN: fabby93



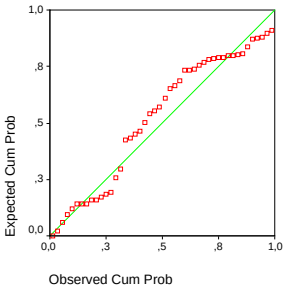
Lognormal P-P Plot of duree inte

LOGIN: fabou



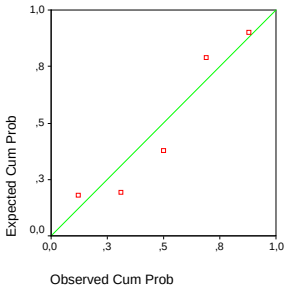
Lognormal P-P Plot of duree inte

LOGIN: fallouis



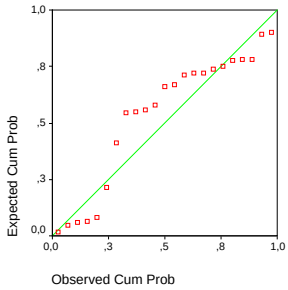
Lognormal P-P Plot of duree inte

LOGIN: fdrugeat



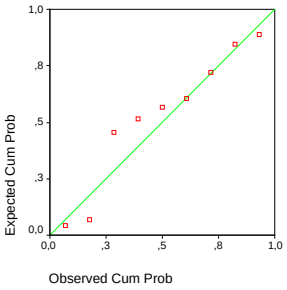
Lognormal P-P Plot of duree inte

LOGIN: fevrard



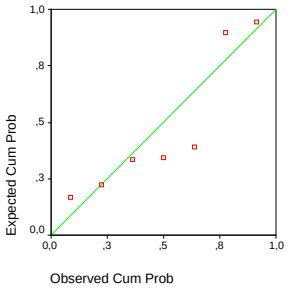
Lognormal P-P Plot of duree inte

LOGIN: FOURNIAL



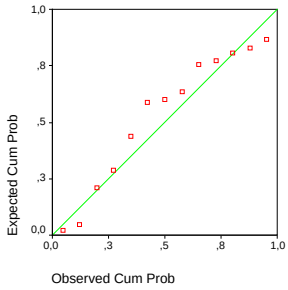
Lognormal P-P Plot of duree inte

LOGIN: Fruor



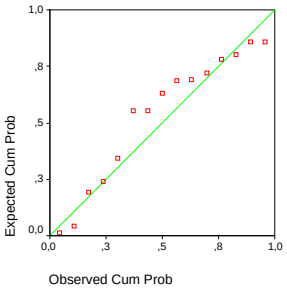
Lognormal P-P Plot of duree inte

LOGIN: frzpower



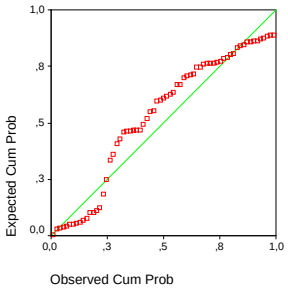
Lognormal P-P Plot of duree inte

LOGIN: GaëlleM



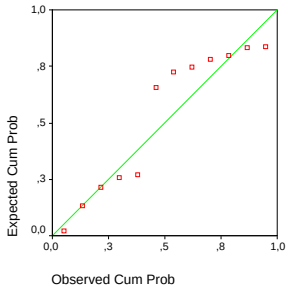
Lognormal P-P Plot of duree inte

LOGIN: gallouin



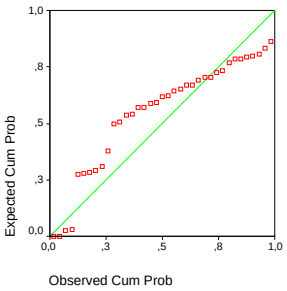
Lognormal P-P Plot of duree inte

LOGIN: Ginie31



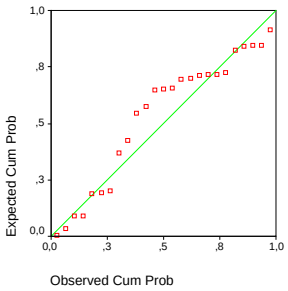
Lognormal P-P Plot of duree inte

LOGIN: gougay



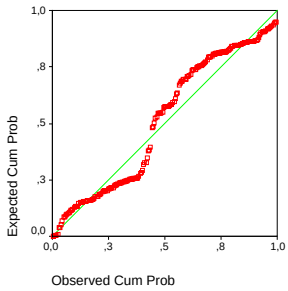
Lognormal P-P Plot of duree inte

LOGIN: Greensky



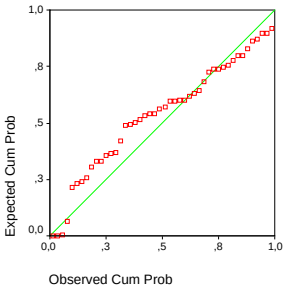
Lognormal P-P Plot of duree inte

LOGIN: guix



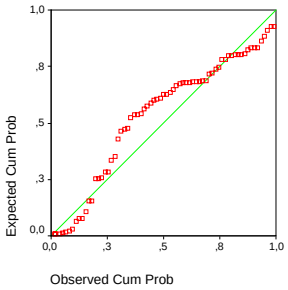
Lognormal P-P Plot of duree inte

LOGIN: gzing



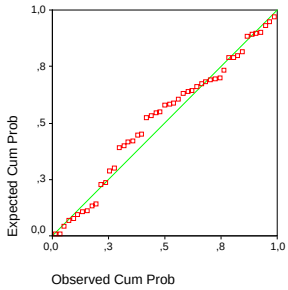
Lognormal P-P Plot of duree inte

LOGIN: hedi.dc



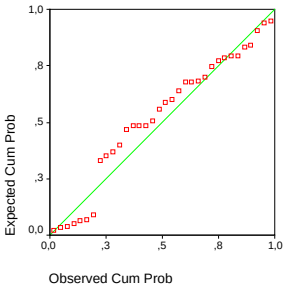
Lognormal P-P Plot of duree inte

LOGIN: HERPIN



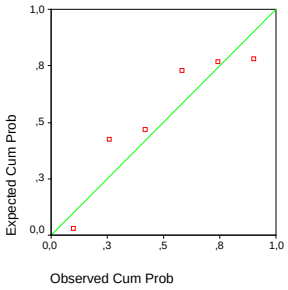
Lognormal P-P Plot of duree inte

LOGIN: hutch



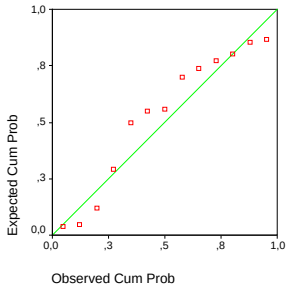
Lognormal P-P Plot of duree inte

LOGIN: j-kolb



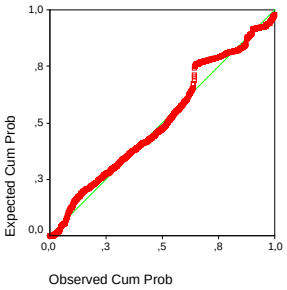
Lognormal P-P Plot of duree inte

LOGIN: jeanbask



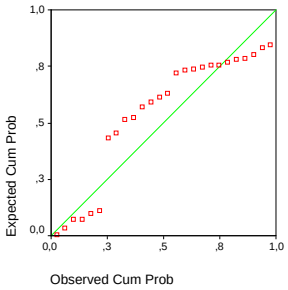
Lognormal P-P Plot of duree inte

LOGIN: jeannot3



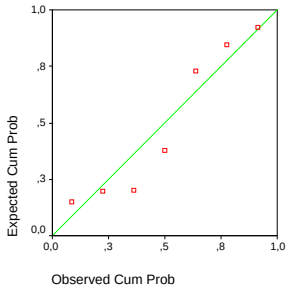
Lognormal P-P Plot of duree inte

LOGIN: jeda



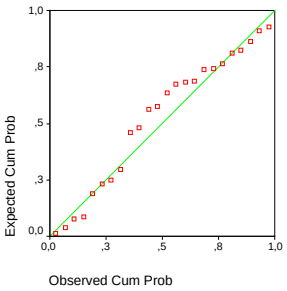
Lognormal P-P Plot of duree inte

LOGIN: JIMASA



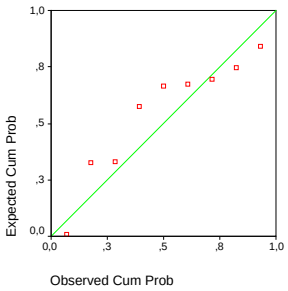
Lognormal P-P Plot of duree inte

LOGIN: joewj



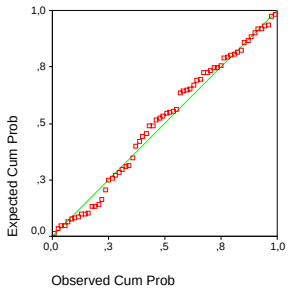
Lognormal P-P Plot of duree inte

LOGIN: josetta



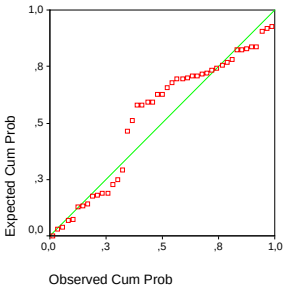
Lognormal P-P Plot of duree inte

LOGIN: juboubou



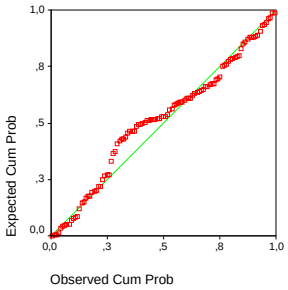
Lognormal P-P Plot of duree inte

LOGIN: juju75



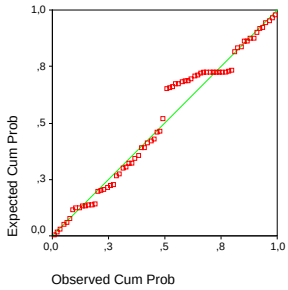
Lognormal P-P Plot of duree inte

LOGIN: kabotine



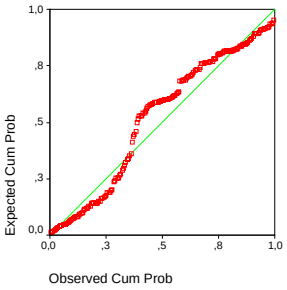
Lognormal P-P Plot of duree inte

LOGIN: Kathrin



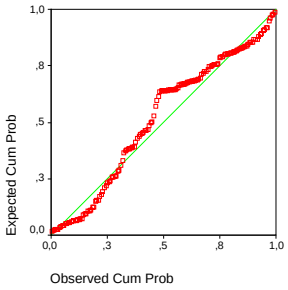
Lognormal P-P Plot of duree inte

LOGIN: Kifran



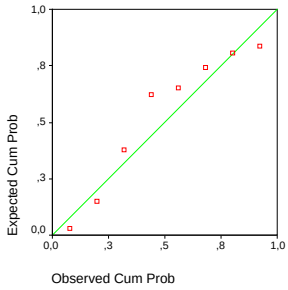
Lognormal P-P Plot of duree inte

LOGIN: lbarbier



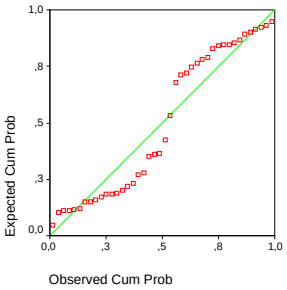
Lognormal P-P Plot of duree inte

LOGIN: leleu5



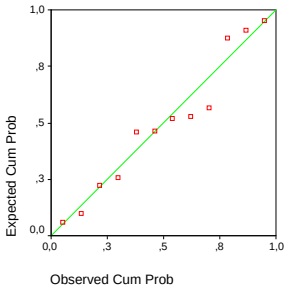
Lognormal P-P Plot of duree inte

LOGIN: levieux



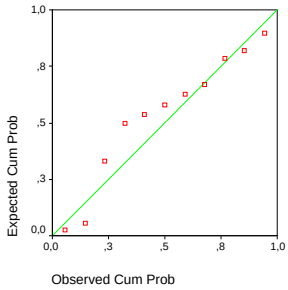
Lognormal P-P Plot of duree inte

LOGIN: lilimimi



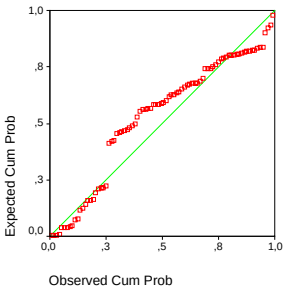
Lognormal P-P Plot of duree inte

LOGIN: LoïcT



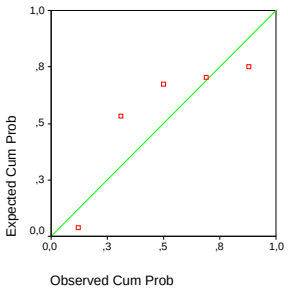
Lognormal P-P Plot of duree inte

LOGIN: Loquet



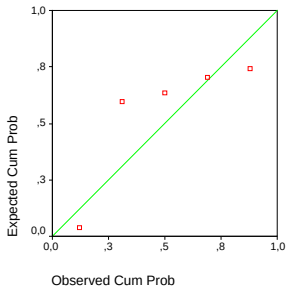
Lognormal P-P Plot of duree inte

LOGIN: luberdov



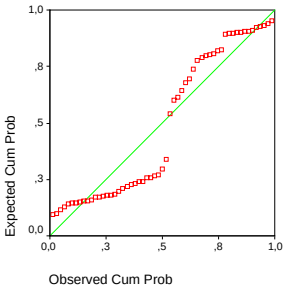
Lognormal P-P Plot of duree inte

LOGIN: ludovicb



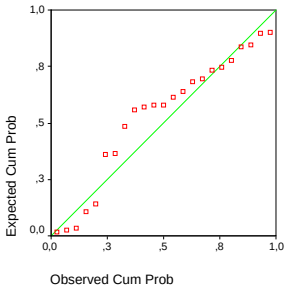
Lognormal P-P Plot of duree inte

LOGIN: m@hy



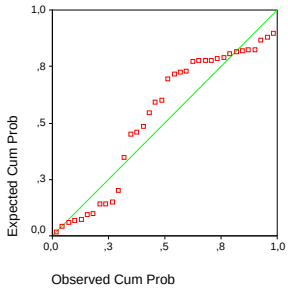
Lognormal P-P Plot of duree inte

LOGIN: manwe



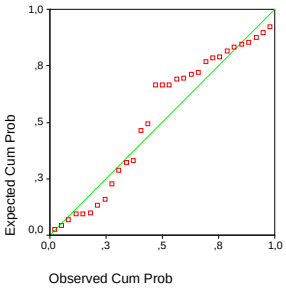
Lognormal P-P Plot of duree inte

LOGIN: mary_ce



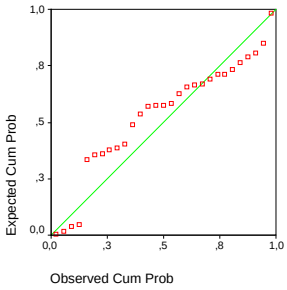
Lognormal P-P Plot of duree inte

LOGIN: Mayfair



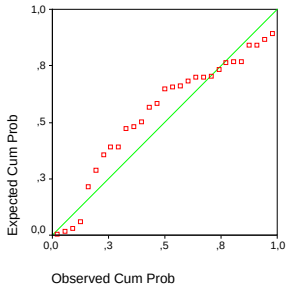
Lognormal P-P Plot of duree inte

LOGIN: mekiela



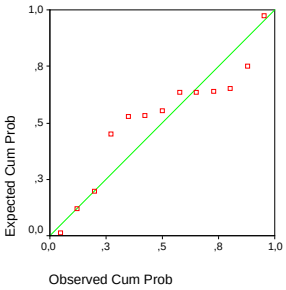
Lognormal P-P Plot of duree inte

LOGIN: Michel



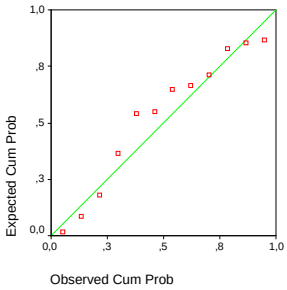
Lognormal P-P Plot of duree inte

LOGIN: micka74



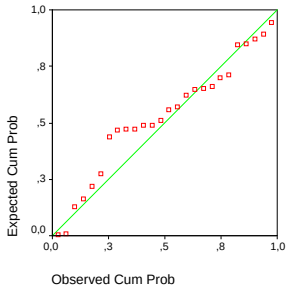
Lognormal P-P Plot of duree inte

LOGIN: mike_s



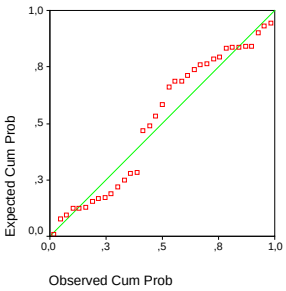
Lognormal P-P Plot of duree inte

LOGIN: miloumoi



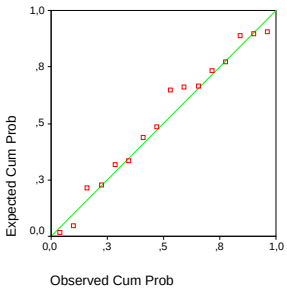
Lognormal P-P Plot of duree inte

LOGIN: minious



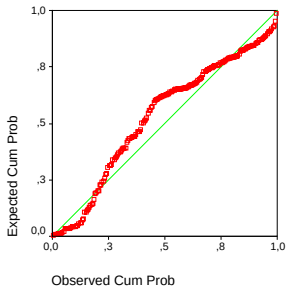
Lognormal P-P Plot of duree inte

LOGIN: modjo-13



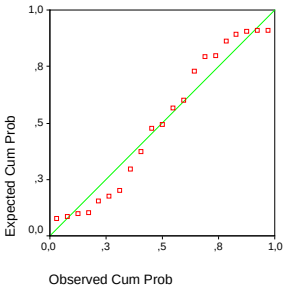
Lognormal P-P Plot of duree inte

LOGIN: mperrot



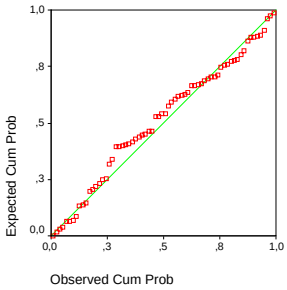
Lognormal P-P Plot of duree inte

LOGIN: Munier



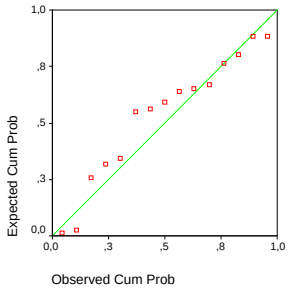
Lognormal P-P Plot of duree inte

LOGIN: Naemesis



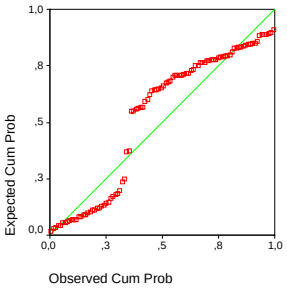
Lognormal P-P Plot of duree inte

LOGIN: Nathalie B



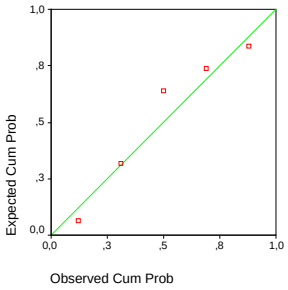
Lognormal P-P Plot of duree inte

LOGIN: nhai07



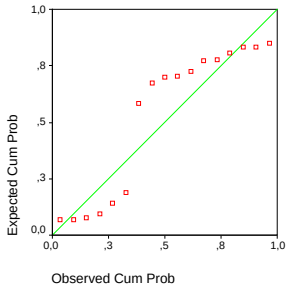
Lognormal P-P Plot of duree inte

LOGIN: nickus



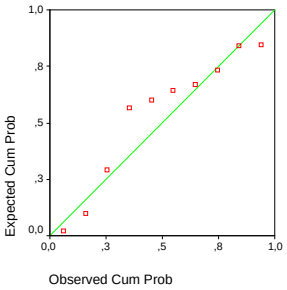
Lognormal P-P Plot of duree inte

LOGIN: ninie12



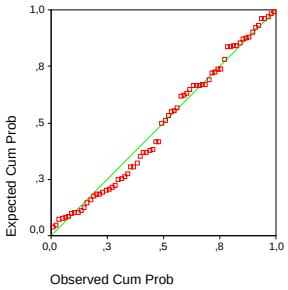
Lognormal P-P Plot of duree inte

LOGIN: noel_net



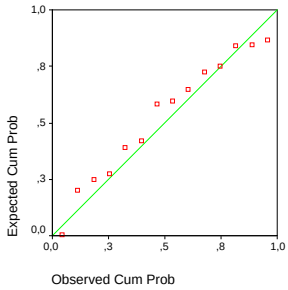
Lognormal P-P Plot of duree inte

LOGIN: philia



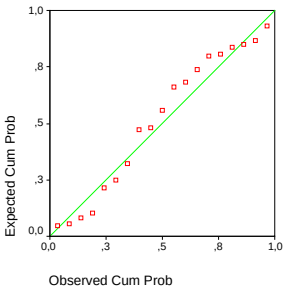
Lognormal P-P Plot of duree inte

LOGIN: picardjf



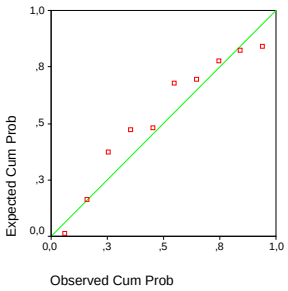
Lognormal P-P Plot of duree inte

LOGIN: popcorn



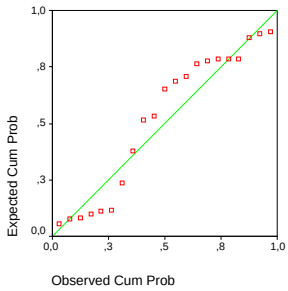
Lognormal P-P Plot of duree inte

LOGIN: poupide



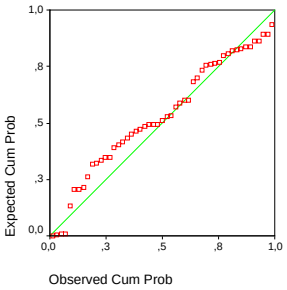
Lognormal P-P Plot of duree inte

LOGIN: prioton



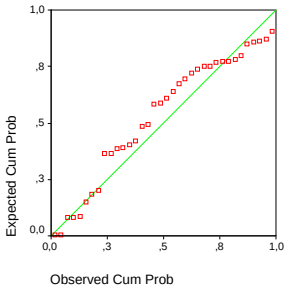
Lognormal P-P Plot of duree inte

LOGIN: prozaiik



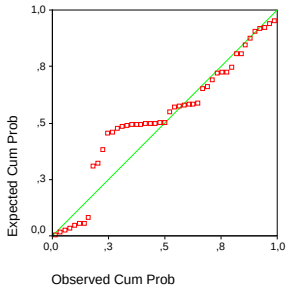
Lognormal P-P Plot of duree inte

LOGIN: quentin



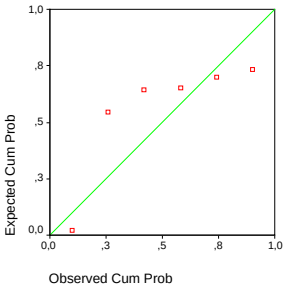
Lognormal P-P Plot of duree inte

LOGIN: RAPH



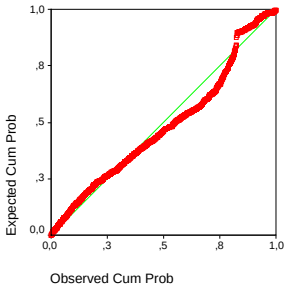
Lognormal P-P Plot of duree inte

LOGIN: relindis



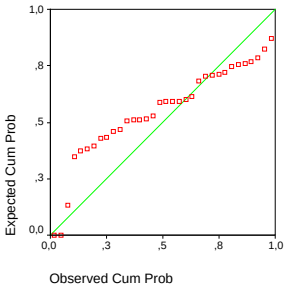
Lognormal P-P Plot of duree inte

LOGIN: rlaval



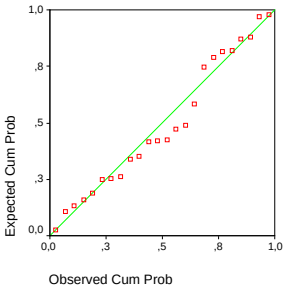
Lognormal P-P Plot of duree inte

LOGIN: romain37



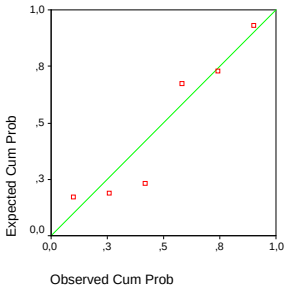
Lognormal P-P Plot of duree inte

LOGIN: sandor



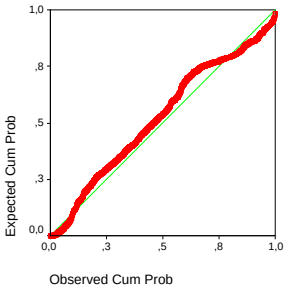
Lognormal P-P Plot of duree inte

LOGIN: sebulba9



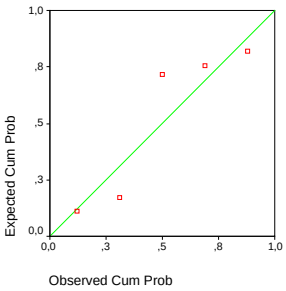
Lognormal P-P Plot of duree inte

LOGIN: semery



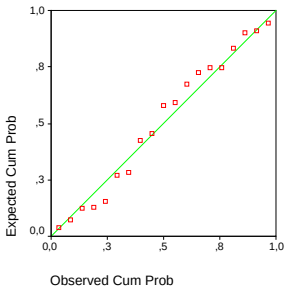
Lognormal P-P Plot of duree inte

LOGIN: sicart



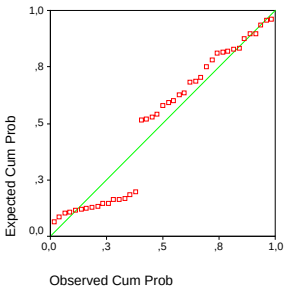
Lognormal P-P Plot of duree inte

LOGIN: sophynet



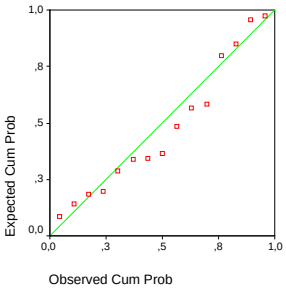
Lognormal P-P Plot of duree inte

LOGIN: Souris



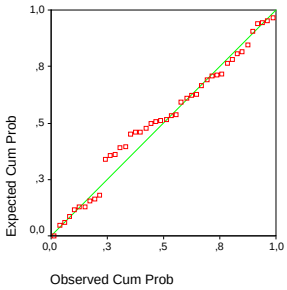
Lognormal P-P Plot of duree inte

LOGIN: starchen



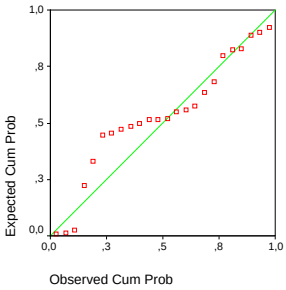
Lognormal P-P Plot of duree inte

LOGIN: Stèph



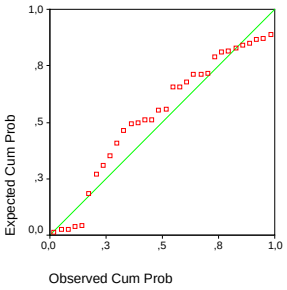
Lognormal P-P Plot of duree inte

LOGIN: stumpf



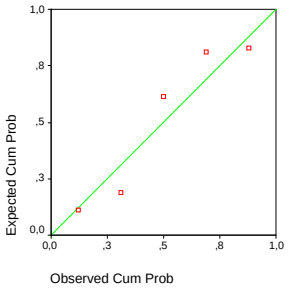
Lognormal P-P Plot of duree inte

LOGIN: Sylvie4



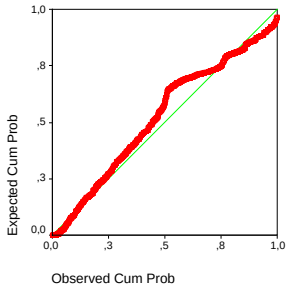
Lognormal P-P Plot of duree inte

LOGIN: Tantan



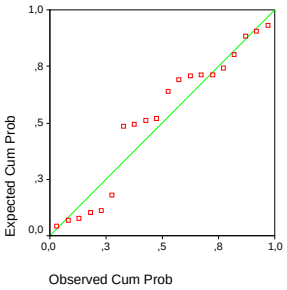
Lognormal P-P Plot of duree inte

LOGIN: Thibaut



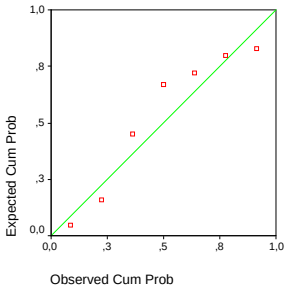
Lognormal P-P Plot of duree inte

LOGIN: Tonio2k



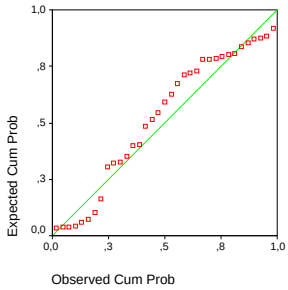
Lognormal P-P Plot of duree inte

LOGIN: toto2



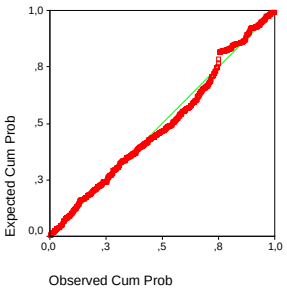
Lognormal P-P Plot of duree inte

LOGIN: tranb



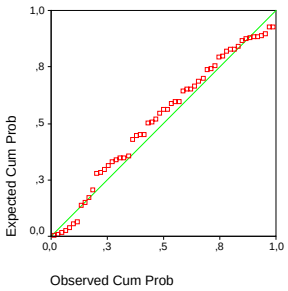
Lognormal P-P Plot of duree inte

LOGIN: vianney2



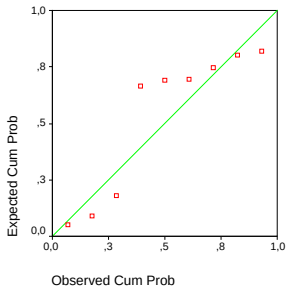
Lognormal P-P Plot of duree inte

LOGIN: Virginie



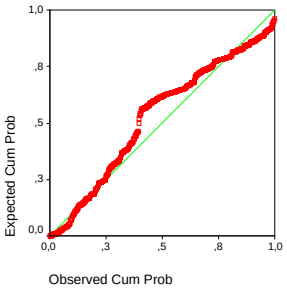
Lognormal P-P Plot of duree inte

LOGIN: xzouf



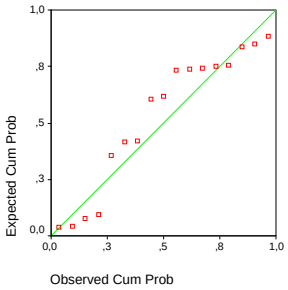
Lognormal P-P Plot of duree inte

LOGIN: Yseulys



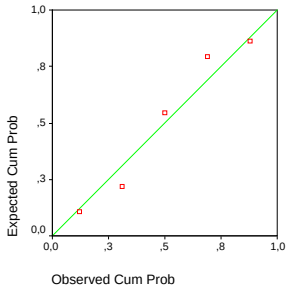
Lognormal P-P Plot of duree inte

LOGIN: Yves

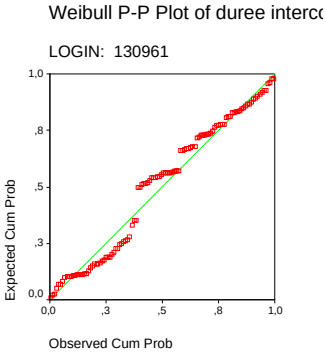


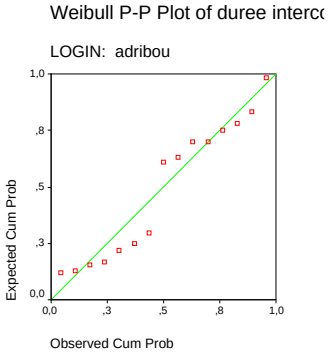
Lognormal P-P Plot of duree inte

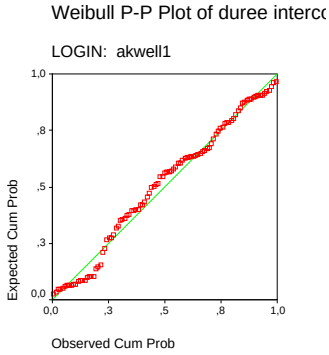
LOGIN: zaultar

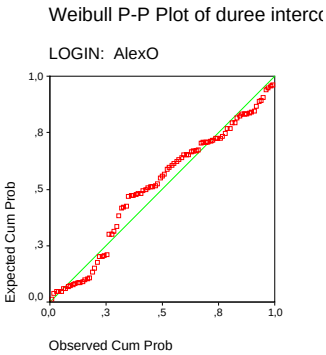


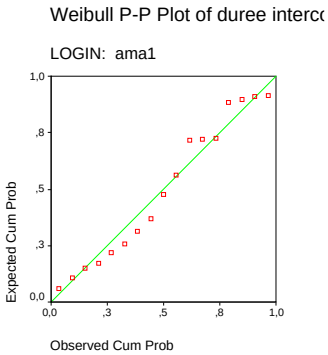
PP PLOT LOI WEIBULL :

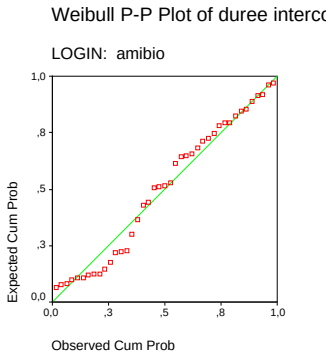


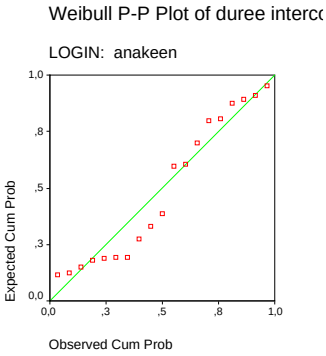


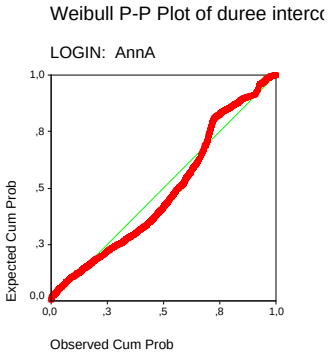


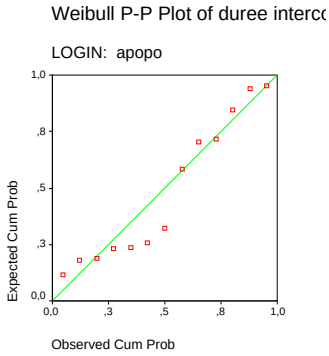


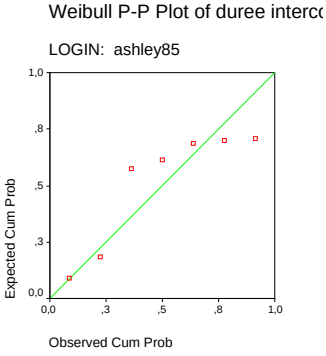


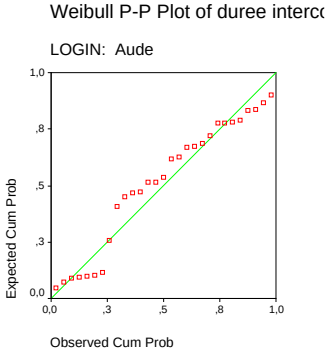


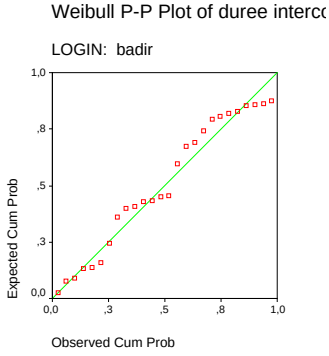






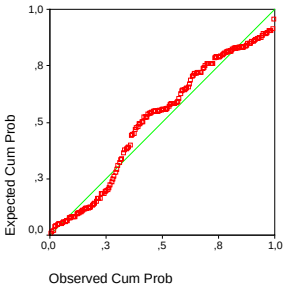






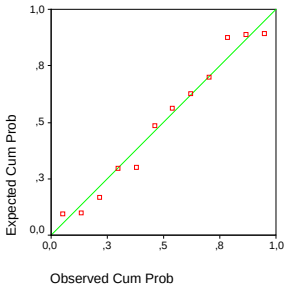
Weibull P-P Plot of duree interc

LOGIN: bgougay



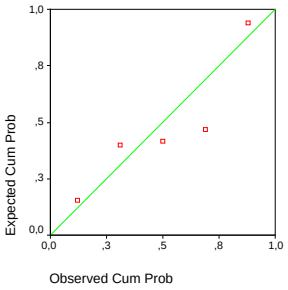
Weibull P-P Plot of duree interc

LOGIN: blondi



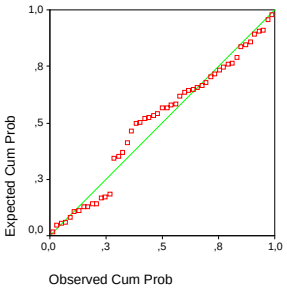
Weibull P-P Plot of duree interc

LOGIN: bontze



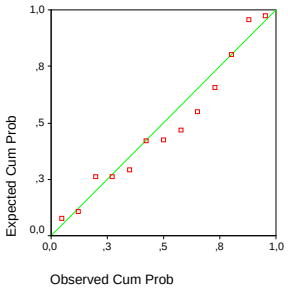
Weibull P-P Plot of duree interc

LOGIN: bpdl



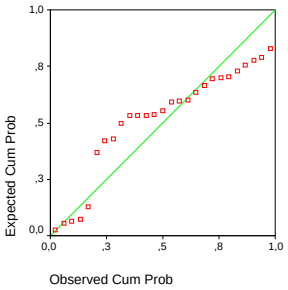
Weibull P-P Plot of duree interc

LOGIN: brusson1



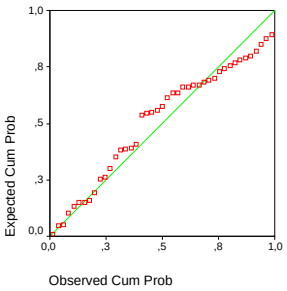
Weibull P-P Plot of duree interc

LOGIN: bruyere



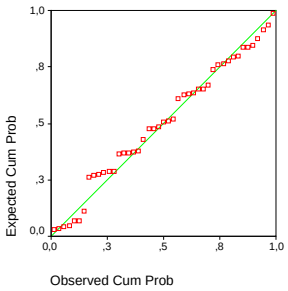
Weibull P-P Plot of duree interc

LOGIN: bschunke



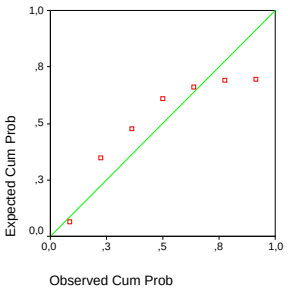
Weibull P-P Plot of duree interc

LOGIN: carolala



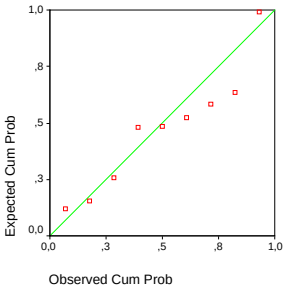
Weibull P-P Plot of duree interc

LOGIN: cathou1



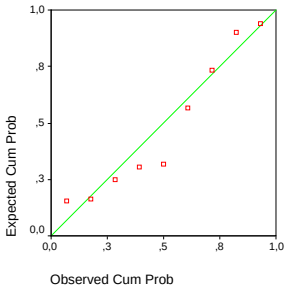
Weibull P-P Plot of duree interc

LOGIN: CATHOU25



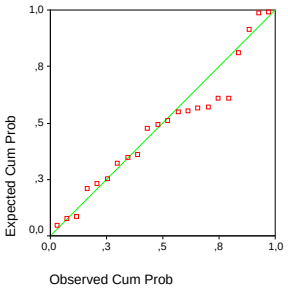
Weibull P-P Plot of duree interc

LOGIN: cclerget



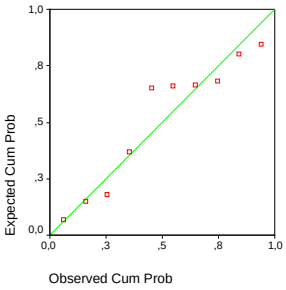
Weibull P-P Plot of duree interc

LOGIN: Céline21



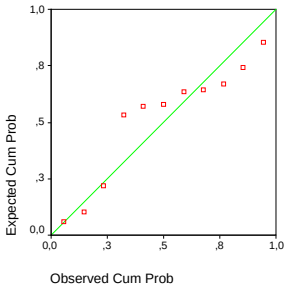
Weibull P-P Plot of duree interc

LOGIN: chap's



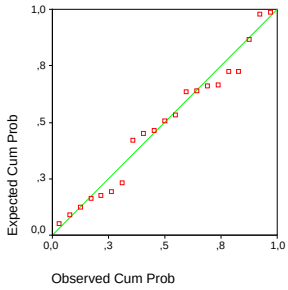
Weibull P-P Plot of duree interc

LOGIN: Chrispi1



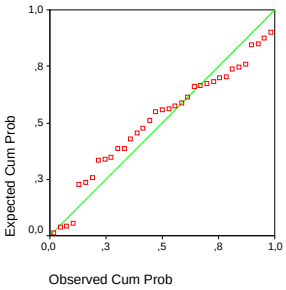
Weibull P-P Plot of duree interc

LOGIN: clemisa



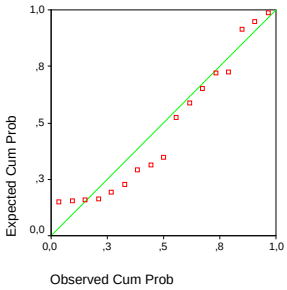
Weibull P-P Plot of duree interc

LOGIN: Cypher



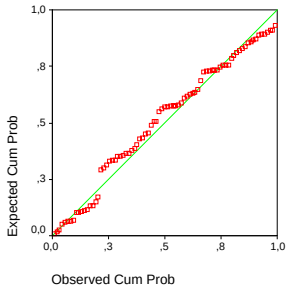
Weibull P-P Plot of duree interc

LOGIN: DALMATE



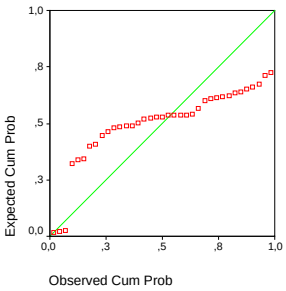
Weibull P-P Plot of duree interc

LOGIN: Delatte



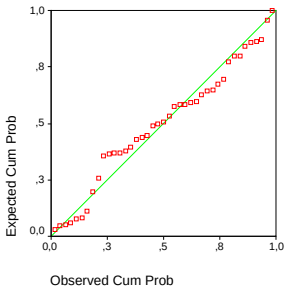
Weibull P-P Plot of duree interc

LOGIN: delillle



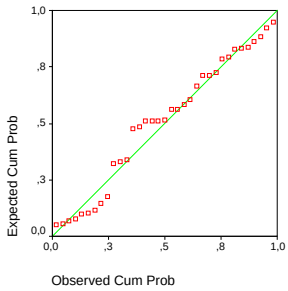
Weibull P-P Plot of duree interc

LOGIN: didine77



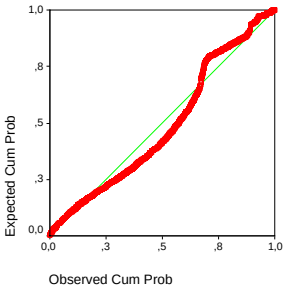
Weibull P-P Plot of duree interc

LOGIN: Dinou



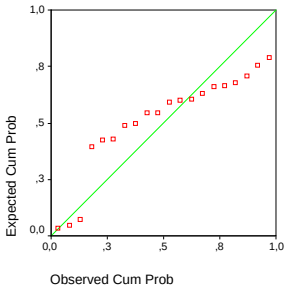
Weibull P-P Plot of duree interc

LOGIN: djoule



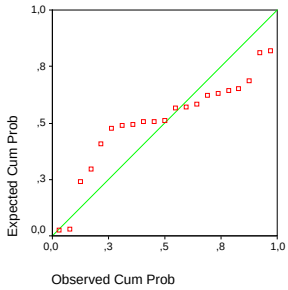
Weibull P-P Plot of duree interc

LOGIN: dolores3



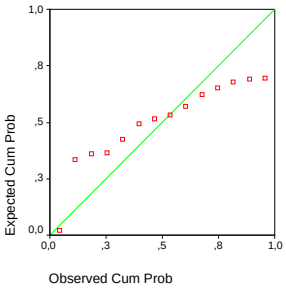
Weibull P-P Plot of duree interc

LOGIN: doppel



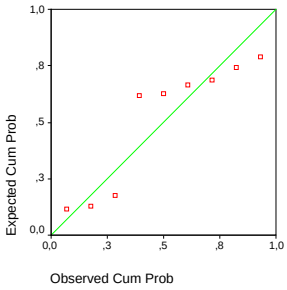
Weibull P-P Plot of duree interc

LOGIN: doppel2



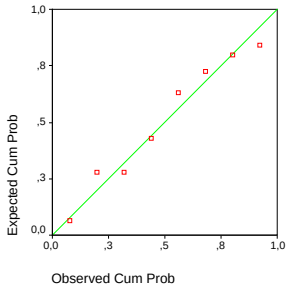
Weibull P-P Plot of duree interc

LOGIN: dperrot



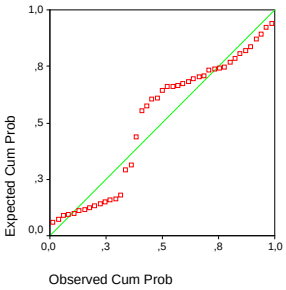
Weibull P-P Plot of duree interc

LOGIN: dzhouari



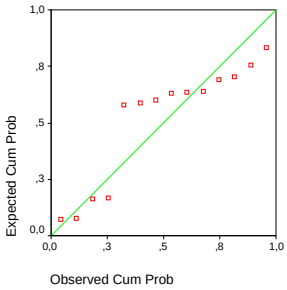
Weibull P-P Plot of duree interc

LOGIN: eb95



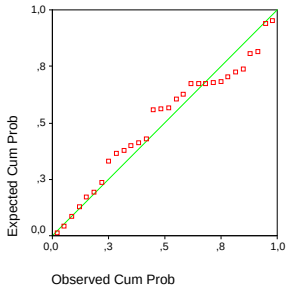
Weibull P-P Plot of duree interc

LOGIN: eb96



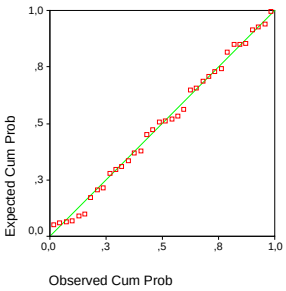
Weibull P-P Plot of duree interc

LOGIN: ehran



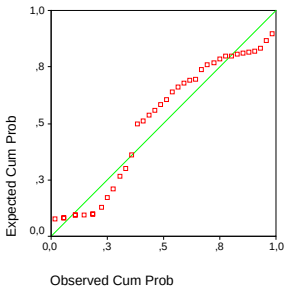
Weibull P-P Plot of duree interc

LOGIN: Emilie



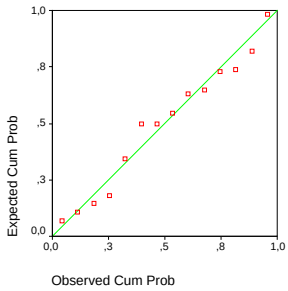
Weibull P-P Plot of duree interc

LOGIN: estia



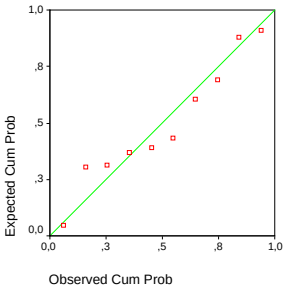
Weibull P-P Plot of duree interc

LOGIN: eusebio1



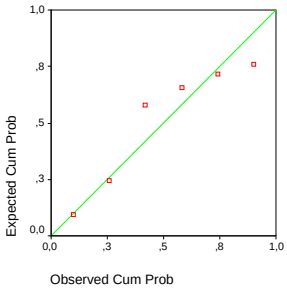
Weibull P-P Plot of duree interc

LOGIN: Fab007



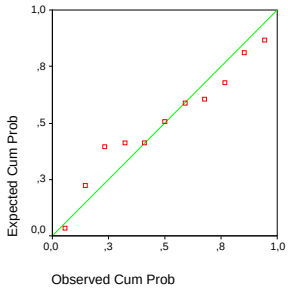
Weibull P-P Plot of duree interc

LOGIN: fabby93



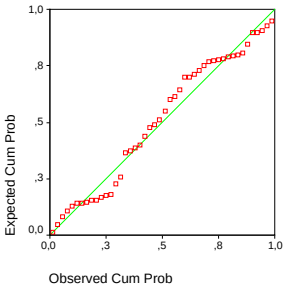
Weibull P-P Plot of duree interc

LOGIN: fabou



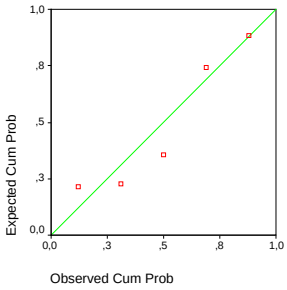
Weibull P-P Plot of duree interc

LOGIN: fallouis



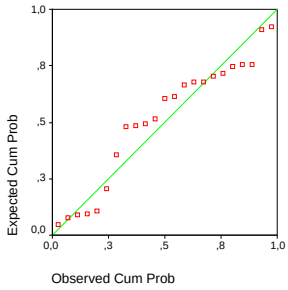
Weibull P-P Plot of duree interc

LOGIN: fdrugeat



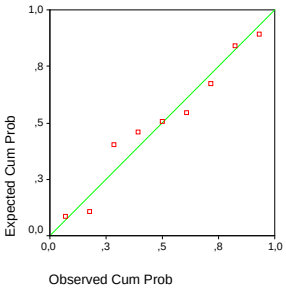
Weibull P-P Plot of duree interc

LOGIN: fevrard



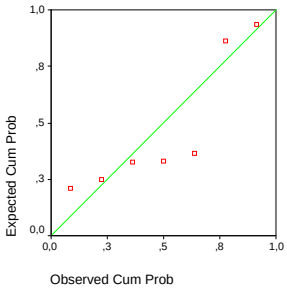
Weibull P-P Plot of duree interc

LOGIN: FOURNIAL



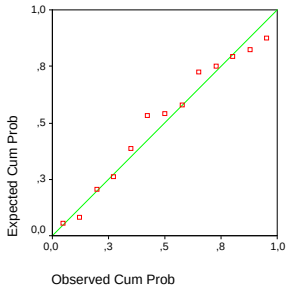
Weibull P-P Plot of duree interc

LOGIN: Fruor



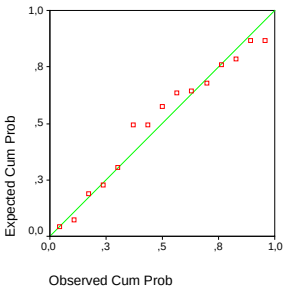
Weibull P-P Plot of duree interc

LOGIN: frzpower



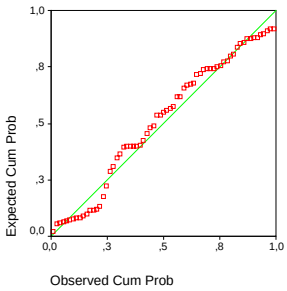
Weibull P-P Plot of duree interc

LOGIN: GaëlleM



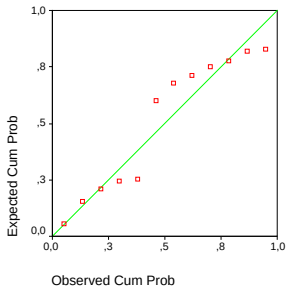
Weibull P-P Plot of duree interc

LOGIN: gallouin



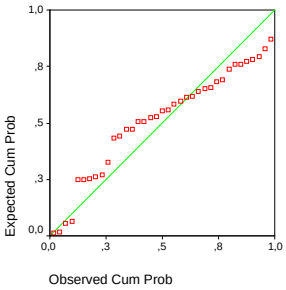
Weibull P-P Plot of duree interc

LOGIN: Ginie31



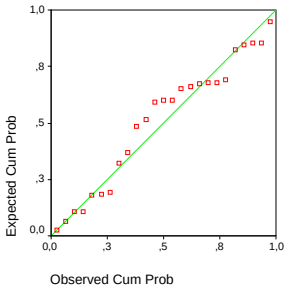
Weibull P-P Plot of duree interc

LOGIN: gougay



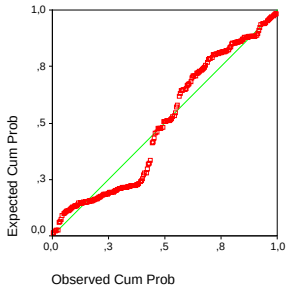
Weibull P-P Plot of duree interc

LOGIN: Greensky



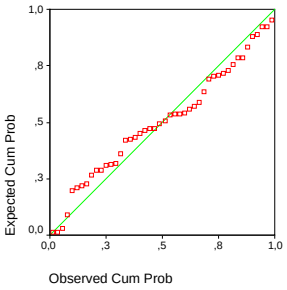
Weibull P-P Plot of duree interc

LOGIN: guix



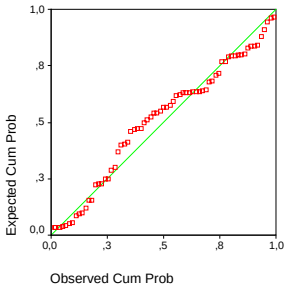
Weibull P-P Plot of duree interc

LOGIN: gzing



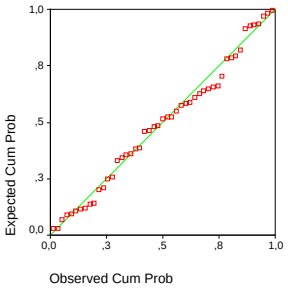
Weibull P-P Plot of duree interc

LOGIN: hedi.dc



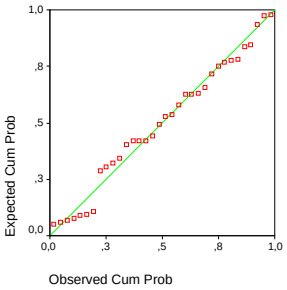
Weibull P-P Plot of duree interc

LOGIN: HERPIN



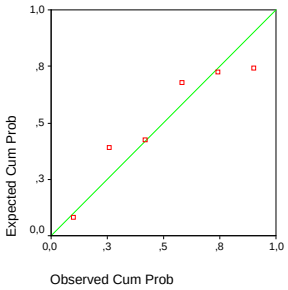
Weibull P-P Plot of duree interc

LOGIN: hutch



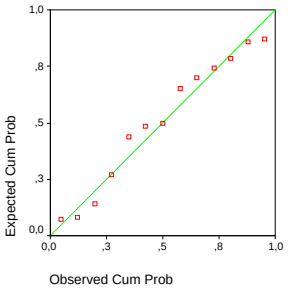
Weibull P-P Plot of duree interc

LOGIN: j-kolb



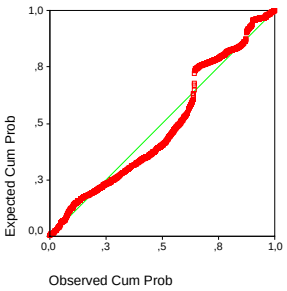
Weibull P-P Plot of duree interc

LOGIN: jeanbask



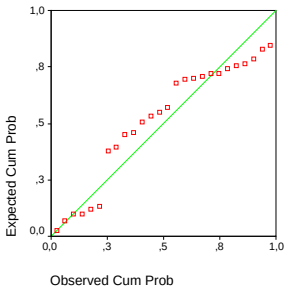
Weibull P-P Plot of duree interc

LOGIN: jeannot3



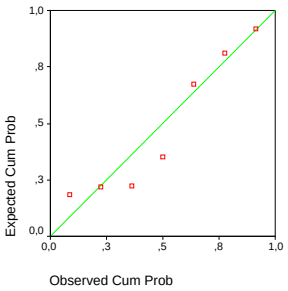
Weibull P-P Plot of duree interc

LOGIN: jeda



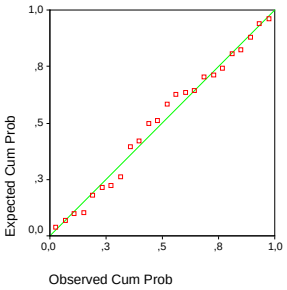
Weibull P-P Plot of duree interc

LOGIN: JIMASA



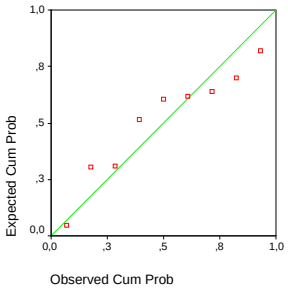
Weibull P-P Plot of duree interc

LOGIN: joewj



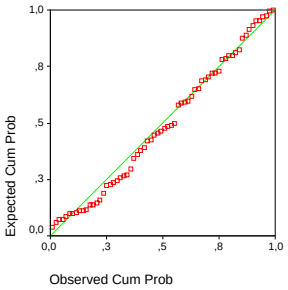
Weibull P-P Plot of duree interc

LOGIN: josetta



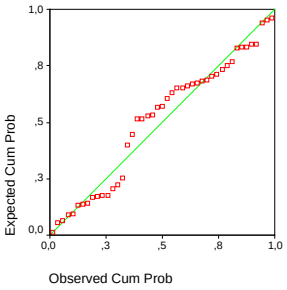
Weibull P-P Plot of duree interc

LOGIN: juboubou



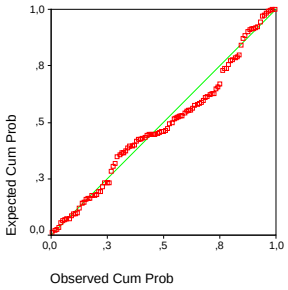
Weibull P-P Plot of duree interc

LOGIN: juju75



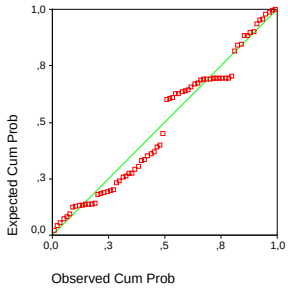
Weibull P-P Plot of duree interc

LOGIN: kabotine



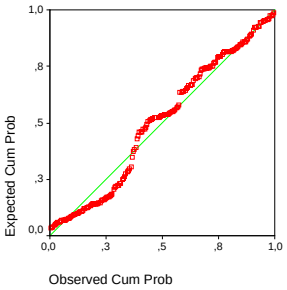
Weibull P-P Plot of duree interc

LOGIN: Kathrin



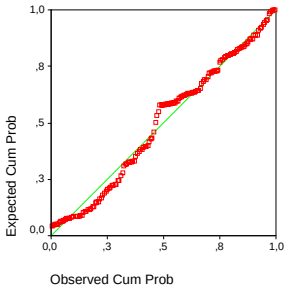
Weibull P-P Plot of duree interc

LOGIN: Kifran



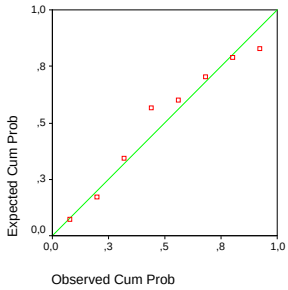
Weibull P-P Plot of duree interc

LOGIN: lbarbier



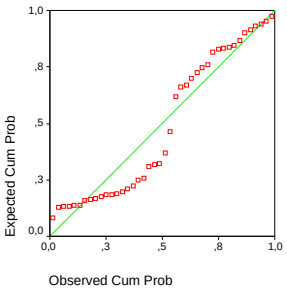
Weibull P-P Plot of duree interc

LOGIN: leleu5



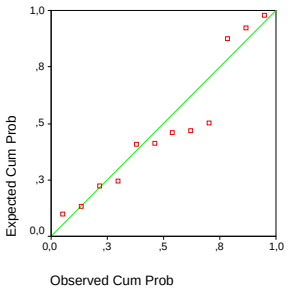
Weibull P-P Plot of duree interc

LOGIN: levieux



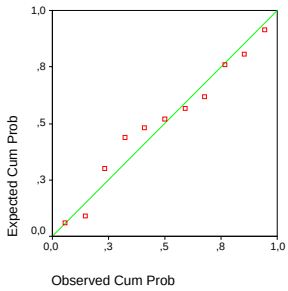
Weibull P-P Plot of duree interc

LOGIN: lilimimi



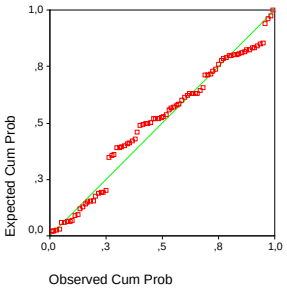
Weibull P-P Plot of duree interc

LOGIN: LoïcT



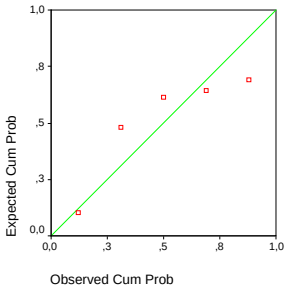
Weibull P-P Plot of duree interc

LOGIN: Loquet



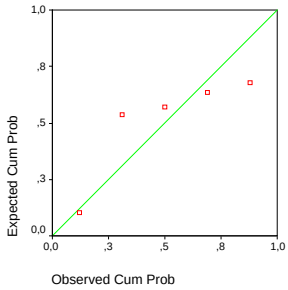
Weibull P-P Plot of duree interc

LOGIN: luberdov



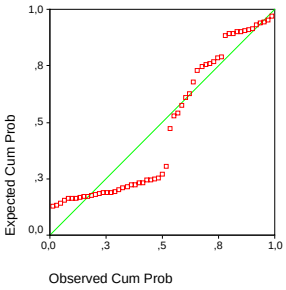
Weibull P-P Plot of duree interc

LOGIN: ludovicb



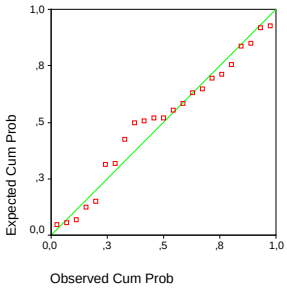
Weibull P-P Plot of duree interc

LOGIN: m@hy



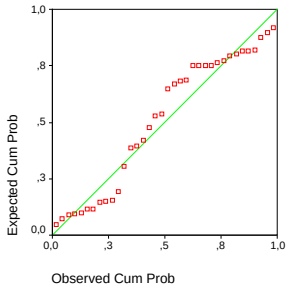
Weibull P-P Plot of duree interc

LOGIN: manwe



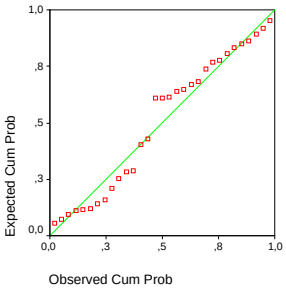
Weibull P-P Plot of duree interc

LOGIN: mary_ce



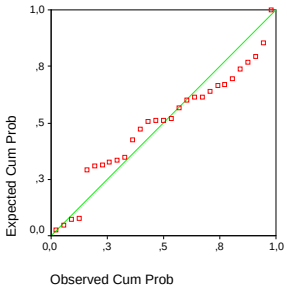
Weibull P-P Plot of duree interc

LOGIN: Mayfair



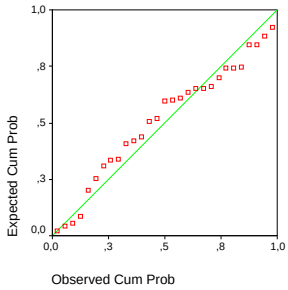
Weibull P-P Plot of duree interc

LOGIN: mekiela



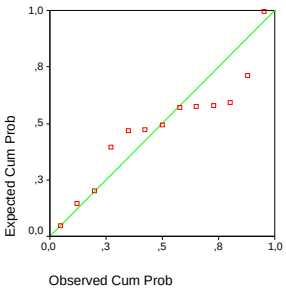
Weibull P-P Plot of duree interc

LOGIN: Michel



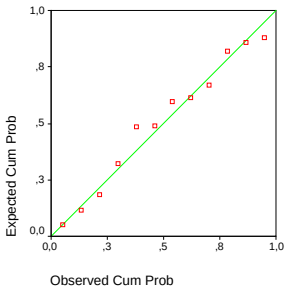
Weibull P-P Plot of duree interc

LOGIN: micka74



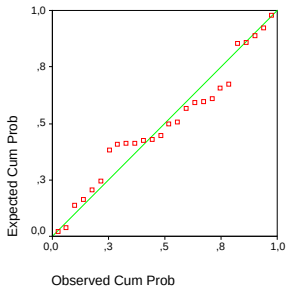
Weibull P-P Plot of duree interc

LOGIN: mike_s



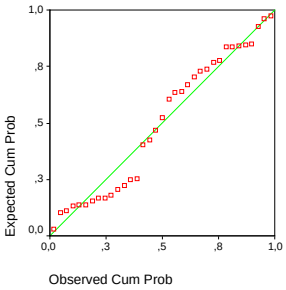
Weibull P-P Plot of duree interc

LOGIN: miloumoi



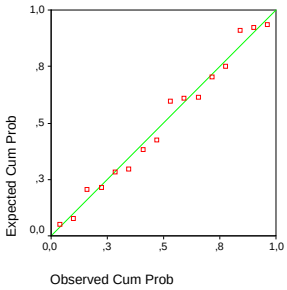
Weibull P-P Plot of duree interc

LOGIN: minious



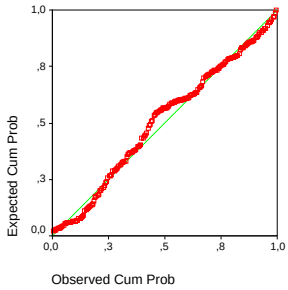
Weibull P-P Plot of duree interc

LOGIN: modjo-13



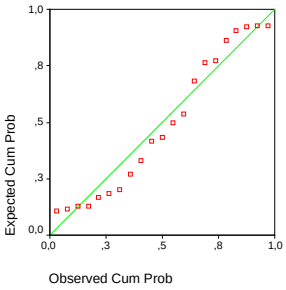
Weibull P-P Plot of duree interc

LOGIN: mperrot



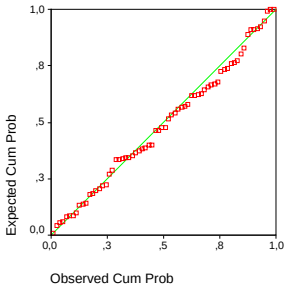
Weibull P-P Plot of duree interc

LOGIN: Munier



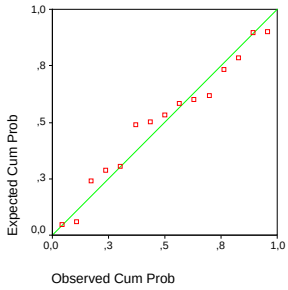
Weibull P-P Plot of duree interc

LOGIN: Naemesis



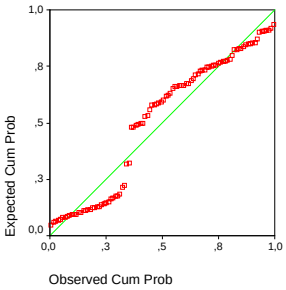
Weibull P-P Plot of duree interc

LOGIN: Nathalie B



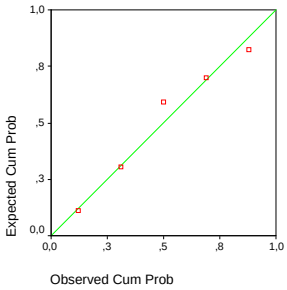
Weibull P-P Plot of duree interc

LOGIN: nhai07



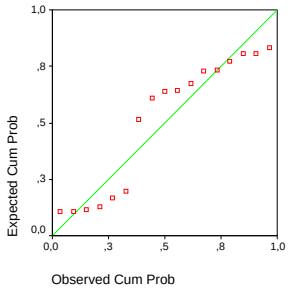
Weibull P-P Plot of duree interc

LOGIN: nickus



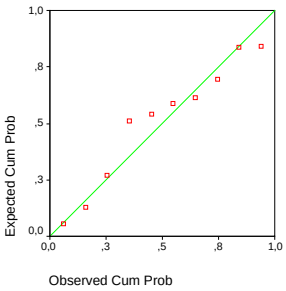
Weibull P-P Plot of duree interc

LOGIN: ninie12



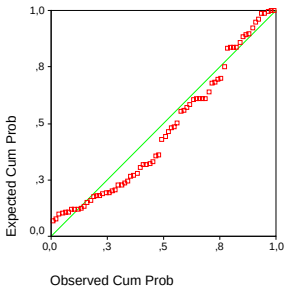
Weibull P-P Plot of duree interc

LOGIN: noel_net



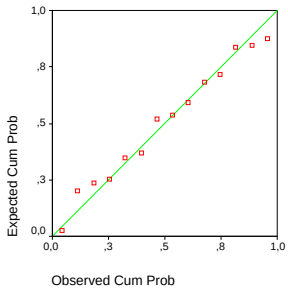
Weibull P-P Plot of duree interc

LOGIN: philia



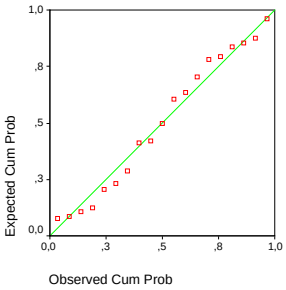
Weibull P-P Plot of duree interc

LOGIN: picardjf



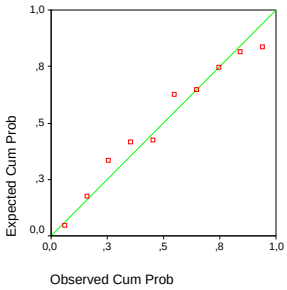
Weibull P-P Plot of duree interc

LOGIN: popcorn



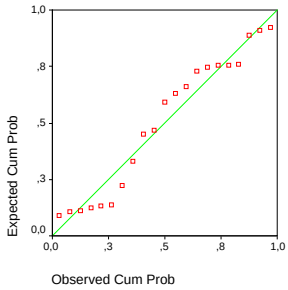
Weibull P-P Plot of duree interc

LOGIN: poupide



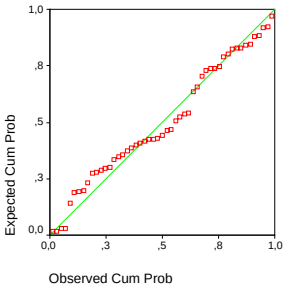
Weibull P-P Plot of duree interc

LOGIN: prioton



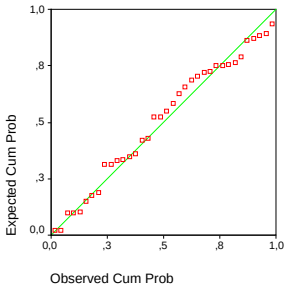
Weibull P-P Plot of duree interc

LOGIN: prozaiik



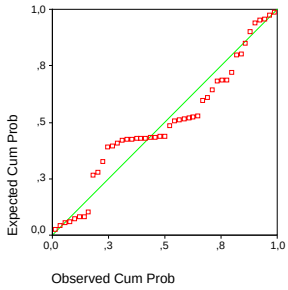
Weibull P-P Plot of duree interc

LOGIN: quentin



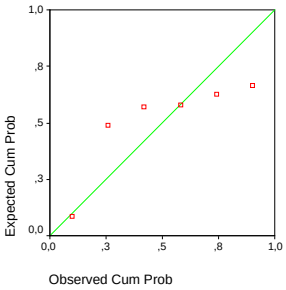
Weibull P-P Plot of duree interc

LOGIN: RAPH



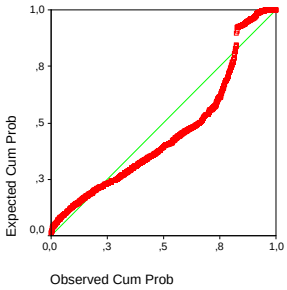
Weibull P-P Plot of duree interc

LOGIN: relindis



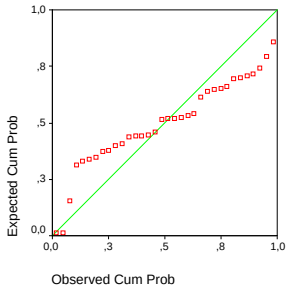
Weibull P-P Plot of duree interc

LOGIN: rlaval



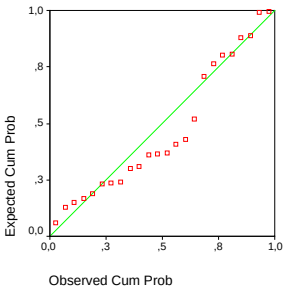
Weibull P-P Plot of duree interc

LOGIN: romain37



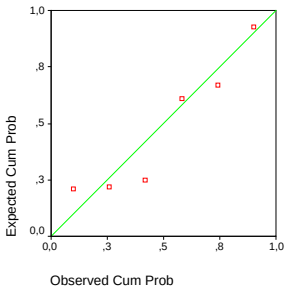
Weibull P-P Plot of duree interc

LOGIN: sandor



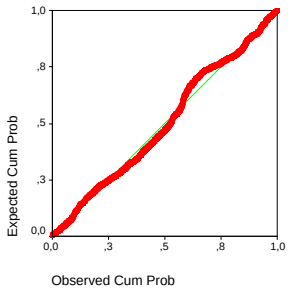
Weibull P-P Plot of duree interc

LOGIN: sebulba9



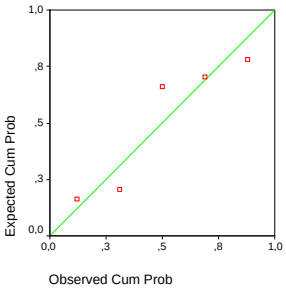
Weibull P-P Plot of duree interc

LOGIN: semery



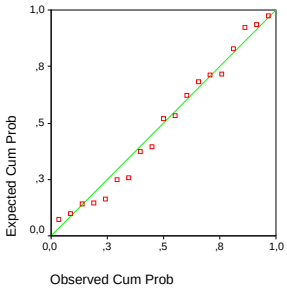
Weibull P-P Plot of duree interc

LOGIN: sicart



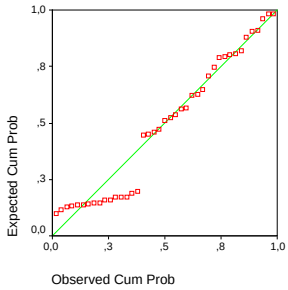
Weibull P-P Plot of duree interc

LOGIN: sophynet



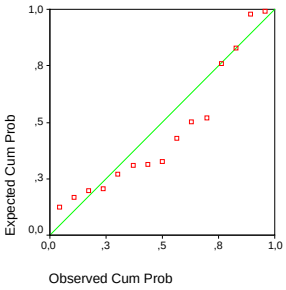
Weibull P-P Plot of duree interc

LOGIN: Souris



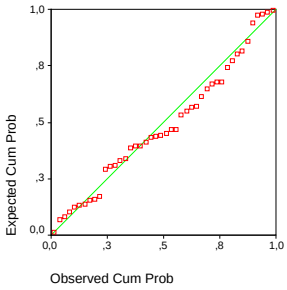
Weibull P-P Plot of duree interc

LOGIN: starchen



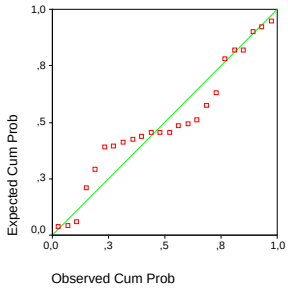
Weibull P-P Plot of duree interc

LOGIN: Stèph



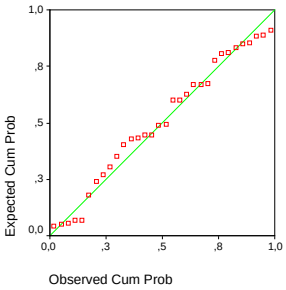
Weibull P-P Plot of duree interc

LOGIN: stumpf



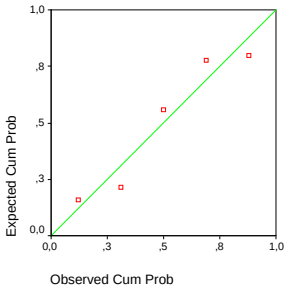
Weibull P-P Plot of duree interc

LOGIN: Sylvie4



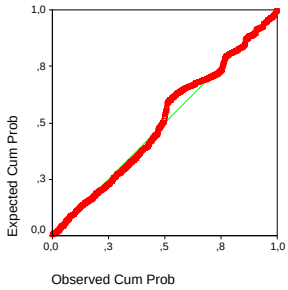
Weibull P-P Plot of duree interc

LOGIN: Tantan



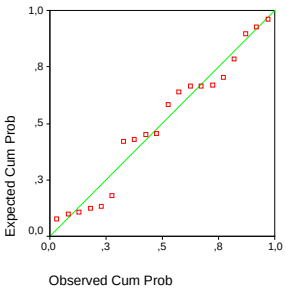
Weibull P-P Plot of duree interc

LOGIN: Thibaut



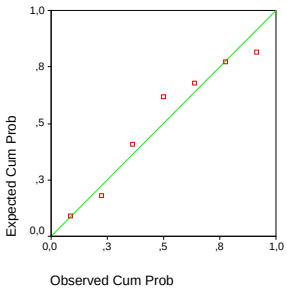
Weibull P-P Plot of duree interc

LOGIN: Tonio2k



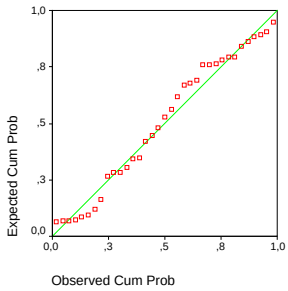
Weibull P-P Plot of duree interc

LOGIN: toto2



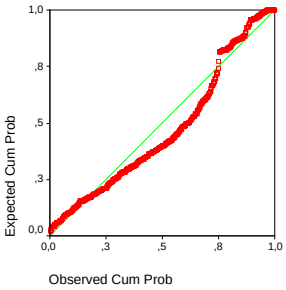
Weibull P-P Plot of duree interc

LOGIN: tranb



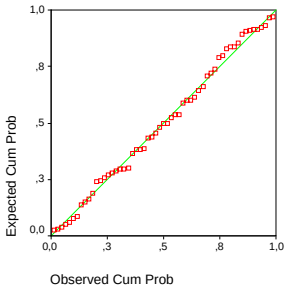
Weibull P-P Plot of duree interc

LOGIN: vianney2



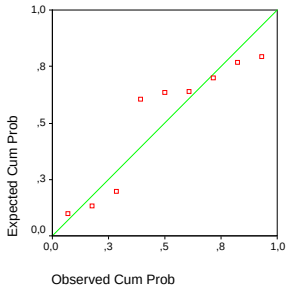
Weibull P-P Plot of duree interc

LOGIN: Virginie



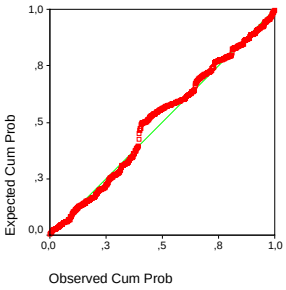
Weibull P-P Plot of duree interc

LOGIN: xzouf



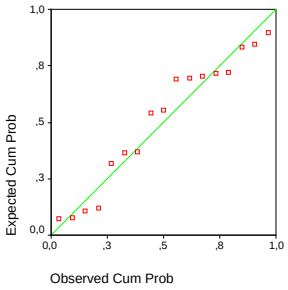
Weibull P-P Plot of duree interc

LOGIN: Yseulys



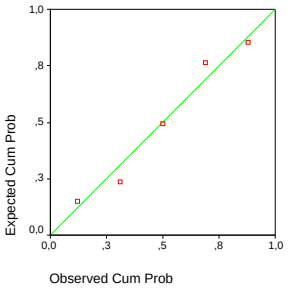
Weibull P-P Plot of duree interc

LOGIN: Yves



Weibull P-P Plot of duree interc

LOGIN: zaultar



ANNEXE 3

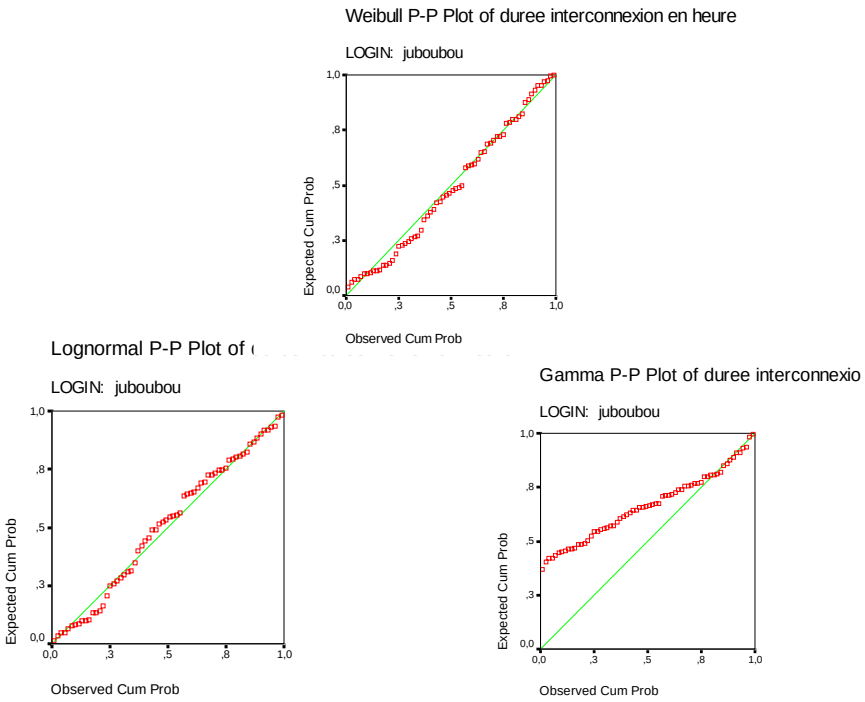
Loi de la durée interconnexion de l'agent A :

Tableau 3: Comparaison des maximums de vraisemblance par loi

Agent A	Gamma	Weibull	Lognormale	Loi choisie
didine77	-104,36	-104,42	-105,8	Gamma ou Weibull
djoule	NCP*	-2221,36	-2369,08	Weibull
juboubou	-144,29	-144,46	-151,22	Gamma ou Weibull
kabotine	-252,86	-251,15	-252,09	Weibull
Loquet	-206,44	-206,51	-210,58	Gamma ou Weibull
m@hy	NCP*	-94,94	-111,24	Weibull
mperrot	-411,62	-411,22	-422,12	Gamma ou Weibull
Naemesis	-172,56	-172,62	-175,11	Gamma ou Weibull
nhai07	-212,6	-213,7	-225,55	Gamma
philia	NCP*	-124,08	-138,02	Weibull
vianney2	NCP*	NCP*	-591,04	Log Normale

Cas juboubou :

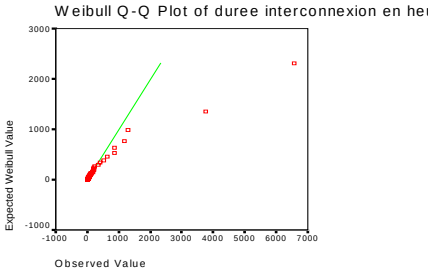
PP Plot : rejet de Gamma (contradictoire avec max. de vraisemblance) et choix de Weibull ou Log normale

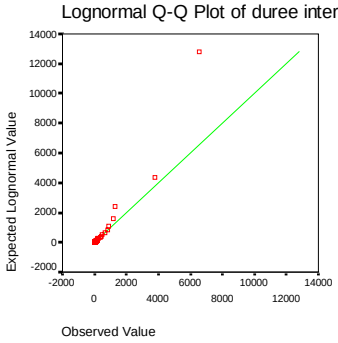


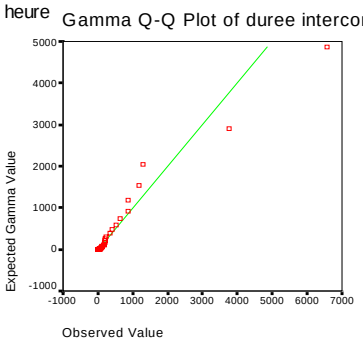
QQ Plot :

Login juboubou
Comparaison des QQ Plots

Nbre d'observ: 66
Excell converge

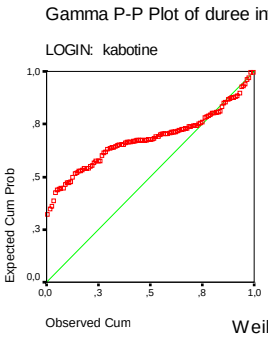


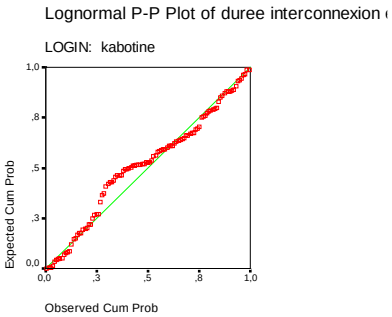


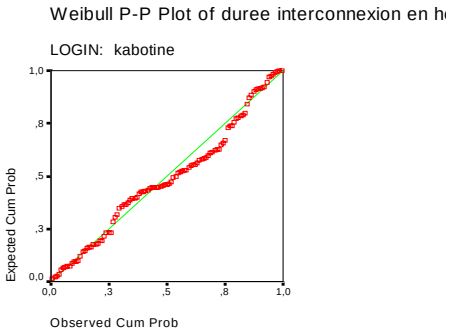


Cas kabotine :

PP Plot :



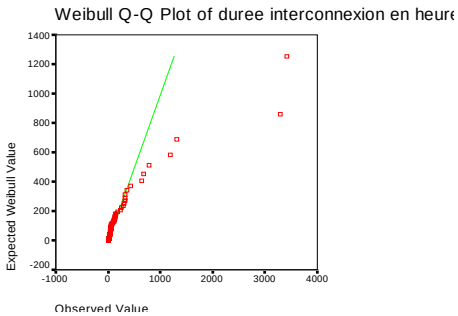


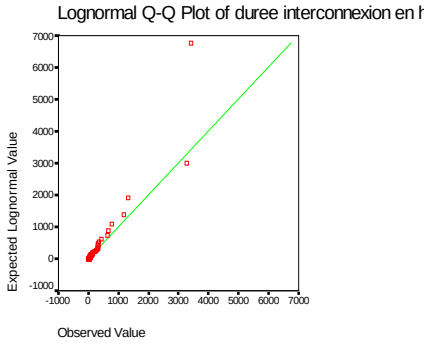


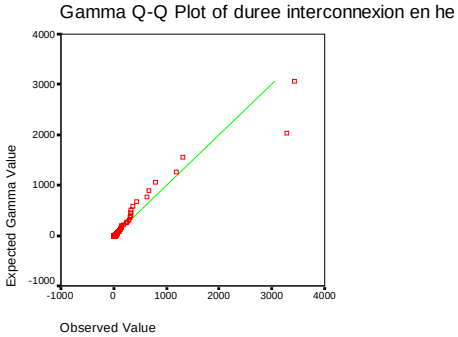
QQ Plot :

Login kabotine
Comparaison des QQ Plots

Nbre d'observ:107
Excell converge

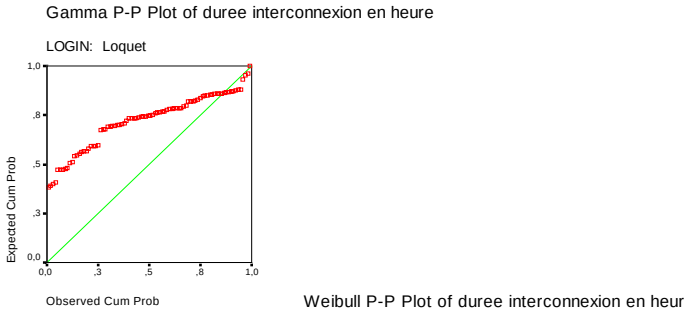


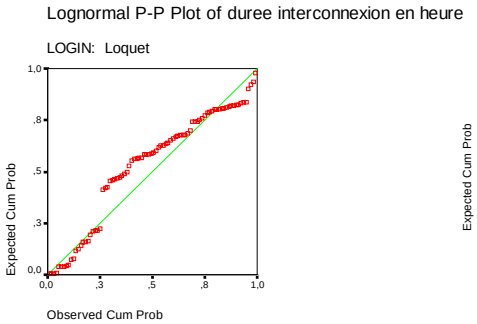


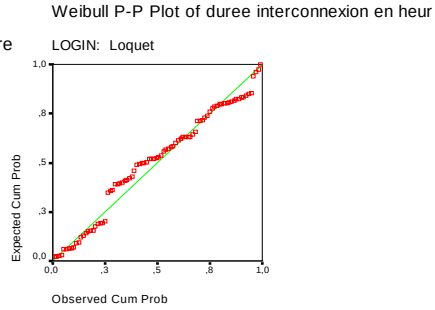


Cas Loquet :

PP plot :



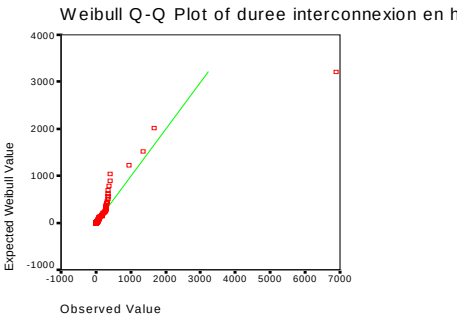


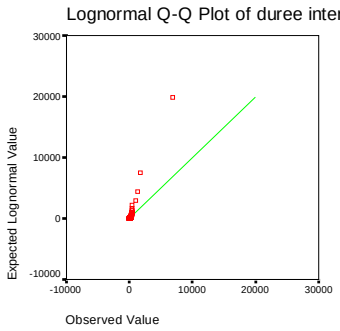


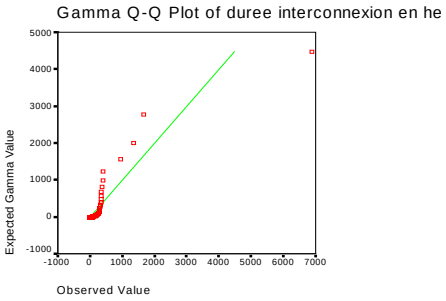
QQ plot :

Login Loquet
Comparaison des QQ Plots

Nbre d'observ:86
Excell converge

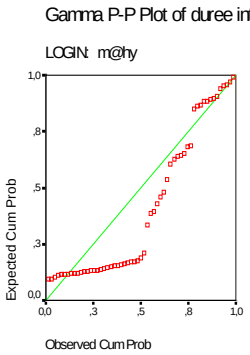


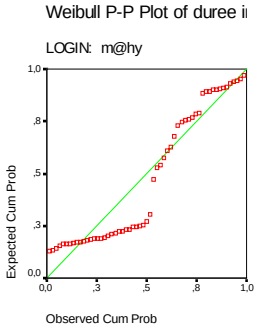


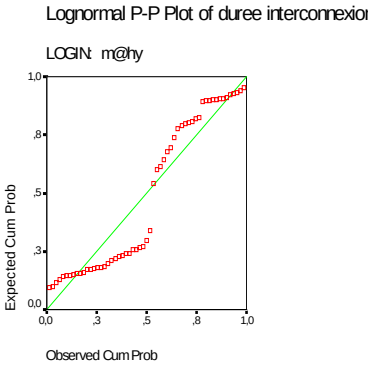


Cas m@hy :

PP plot :



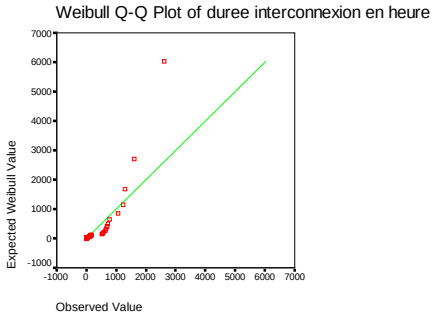


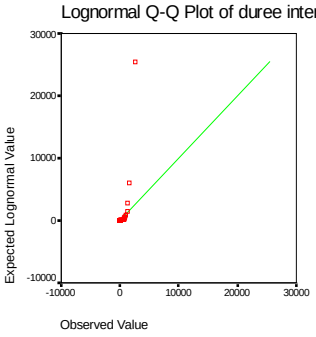


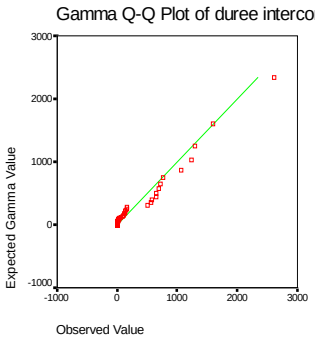
QQ Plot :

Login m@hy
Comparaison des QQ Plots

Nbre d'observ: 57
Excell: seul Gamma
ne converge pas

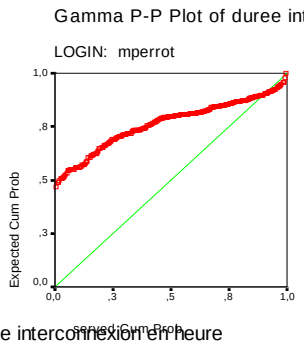


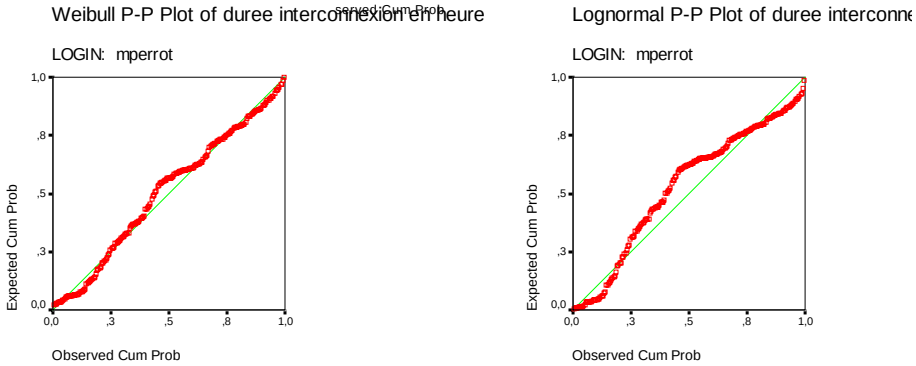




Cas mperrot :

PP plot :

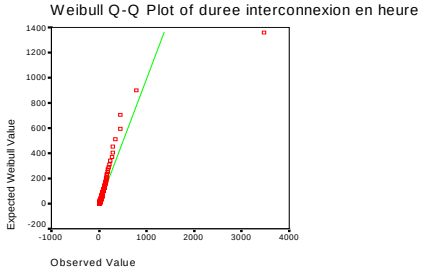


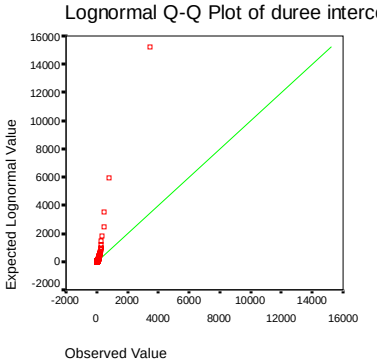


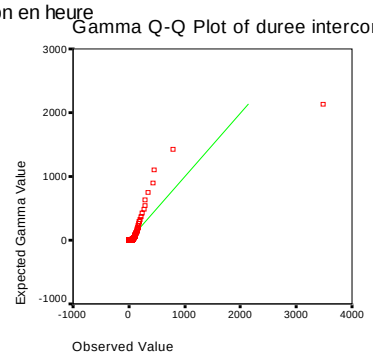
QQ Plot :

Login mperrot
Comparaison des QQ Plots

Nbre d'observ: 191
Excell converge

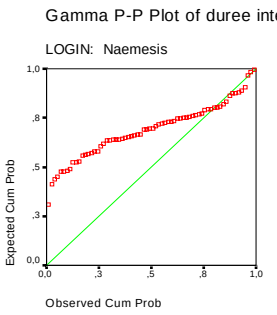


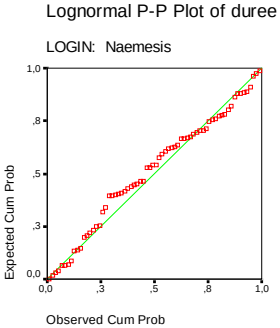


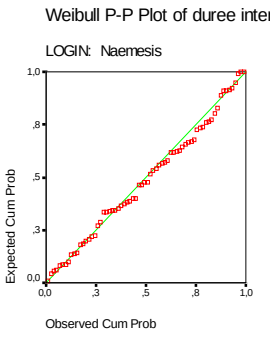


Cas Naemesis :

PP plot :



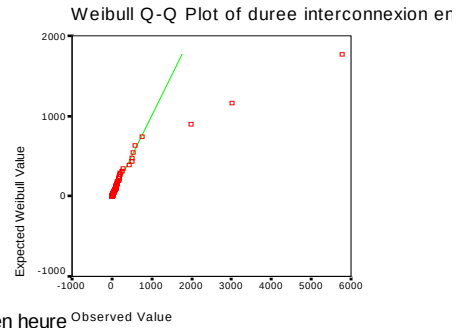


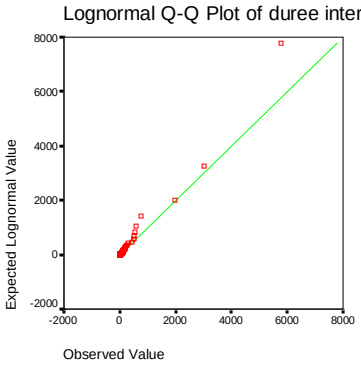


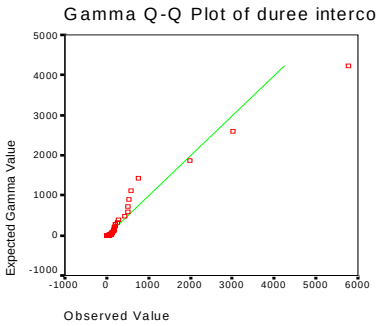
QQ plot :

Login Naemesis
Comparaison des QQ Plots

Nbre d'observ: 68
Excell converge

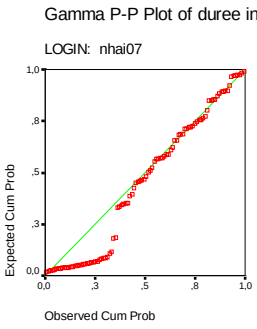


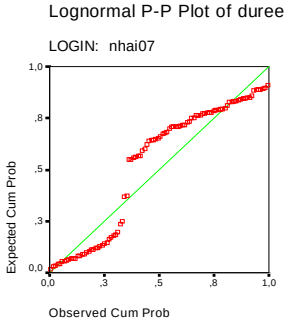


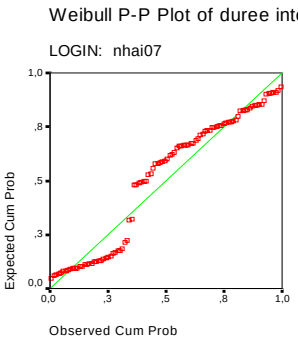


Cas nhai07 :

PP plot :



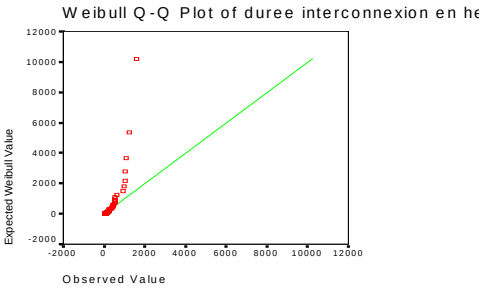


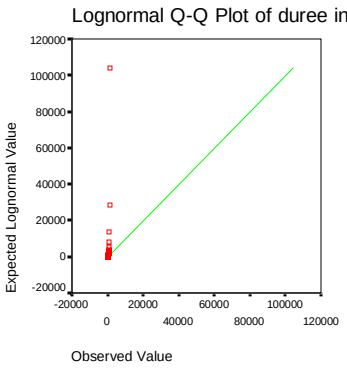


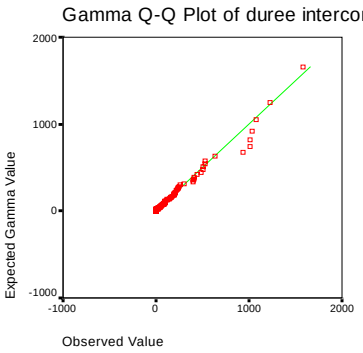
QQ Plot :

Login nhai07
Comparaison des QQ Plots

Nbre d'observ: 98
Excell converge

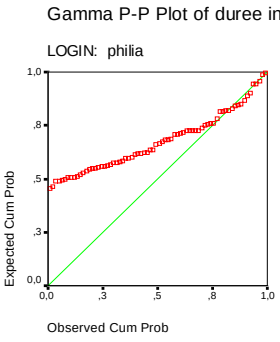


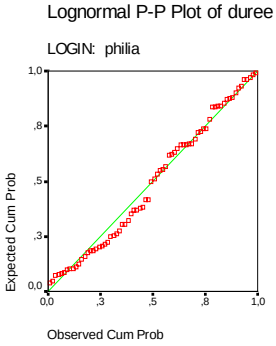


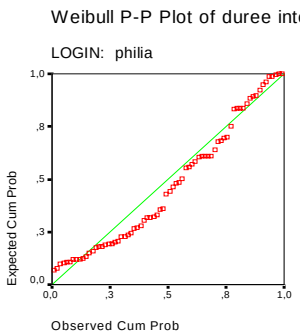


Cas philia :

PP plot :



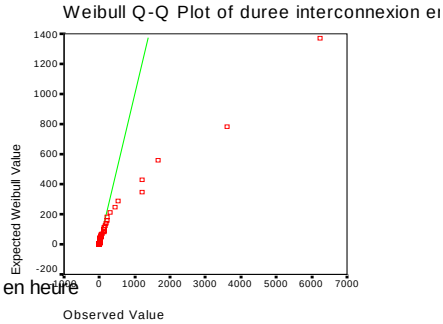


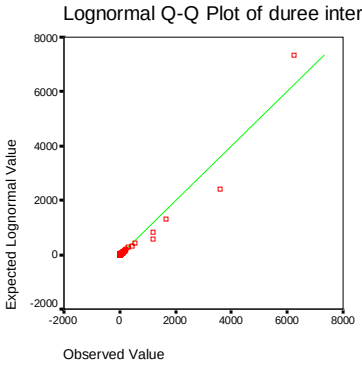


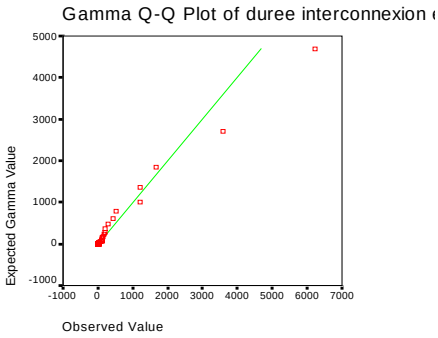
QQ plot :

Login philia
Comparaison des QQ Plots

Nbre d'observ: 72
Excell: Seul gamma
ne converge pas

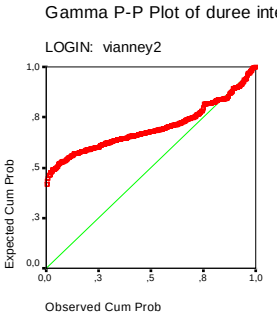


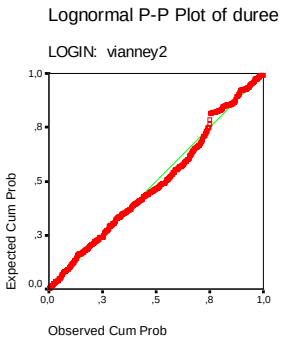


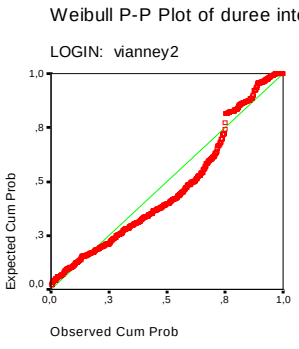


Cas vianney2 :

PP plot :



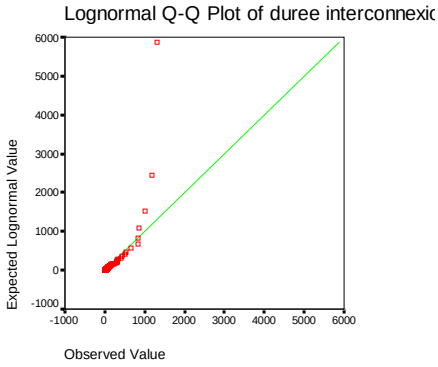


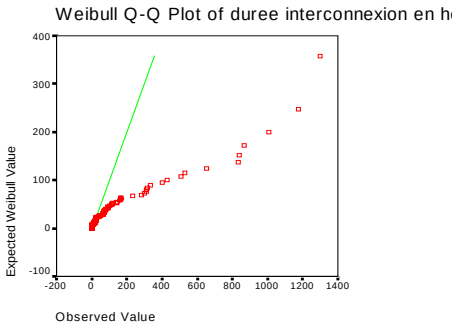


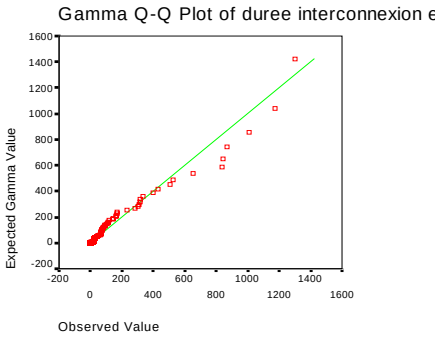
QQ Plot :

Login vianney2
Comparaison des QQ Plots

Nbre d'observ: 425
Excell: Gamma et
Weibull ne convergent pas





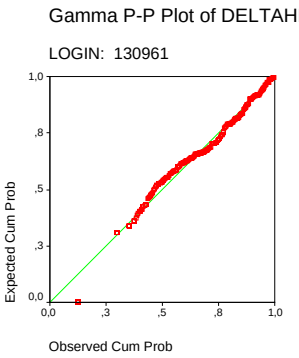


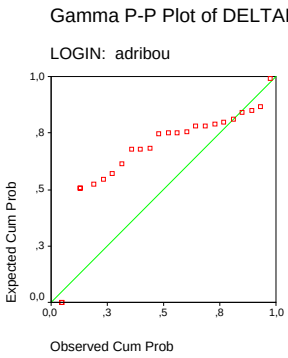
ANNEXE 4

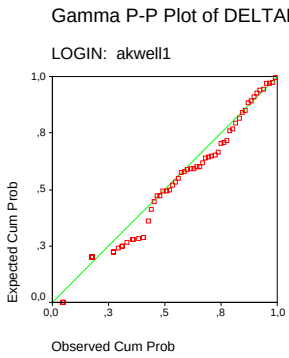
Loi de durée interconnexion des agents B pour chaque login d'un agent A.

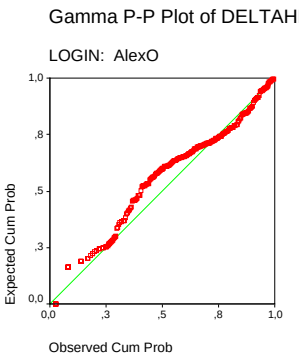
Ces résultats ne concernent que des logins d'agent A pour lesquels des agents B se sont connectés. Il n'y a donc pas 135 cas comme pour l'annexe 2, mais 105 cas présentés ci-dessous (cas où le nombre de connections de B est supérieur à 2).

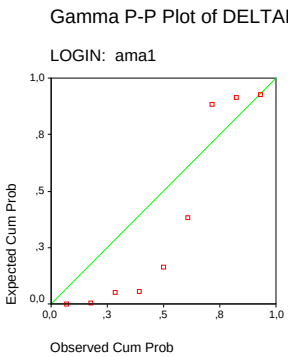
PP PLOT LOI GAMMA:

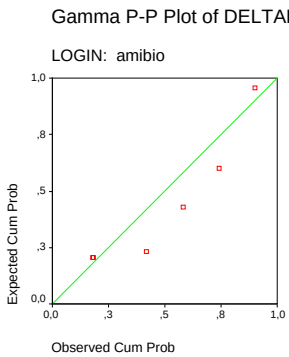


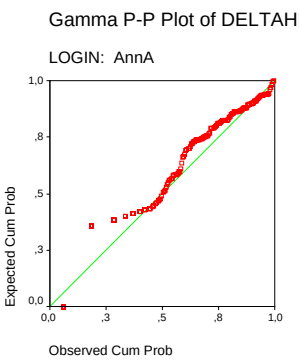


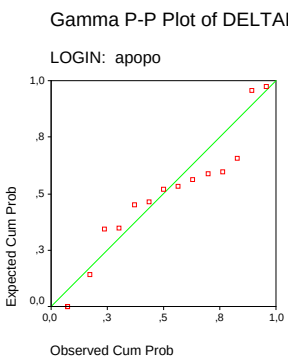


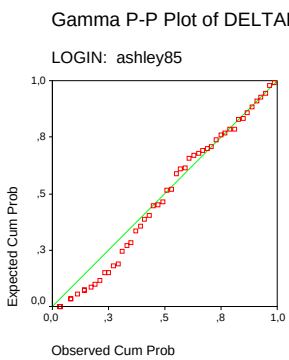






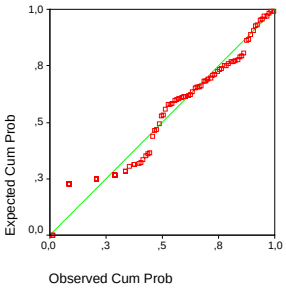






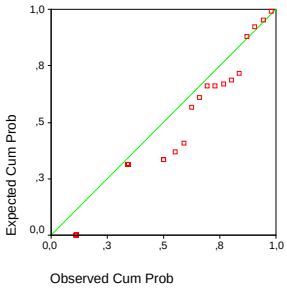
Gamma P-P Plot of DELTAHEU

LOGIN: Aude



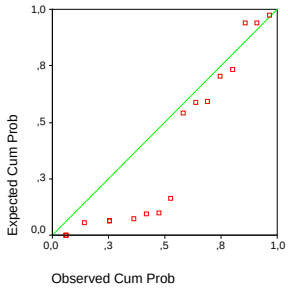
Gamma P-P Plot of DELTAHEU

LOGIN: badir



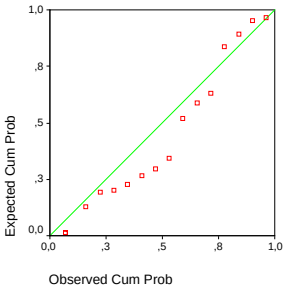
Gamma P-P Plot of DELTAHEU

LOGIN: bgougay



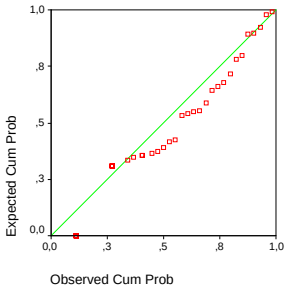
Gamma P-P Plot of DELTAHEU

LOGIN: blondi



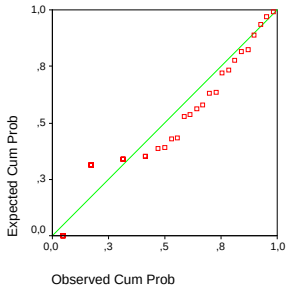
Gamma P-P Plot of DELTAHEU

LOGIN: bontze



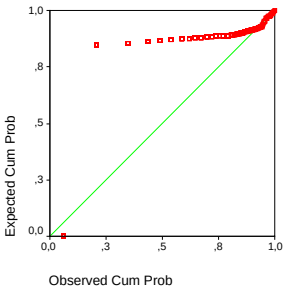
Gamma P-P Plot of DELTAHEU

LOGIN: bpdl



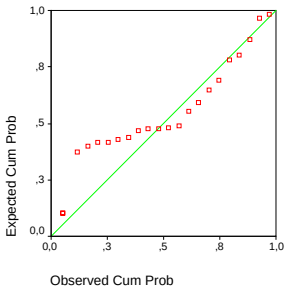
Gamma P-P Plot of DELTAHEU

LOGIN: brusson1



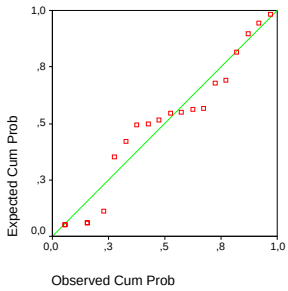
Gamma P-P Plot of DELTAHEU

LOGIN: bruyere



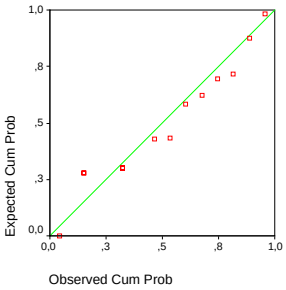
Gamma P-P Plot of DELTAHEU

LOGIN: carolala



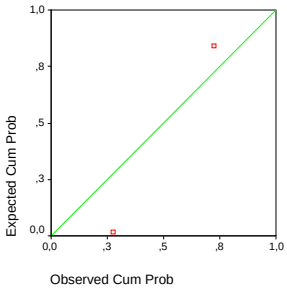
Gamma P-P Plot of DELTAHEU

LOGIN: cathou1



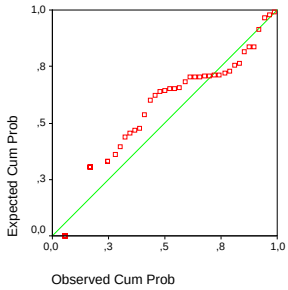
Gamma P-P Plot of DELTAHEU

LOGIN: CATHOU25



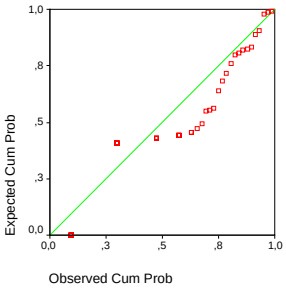
Gamma P-P Plot of DELTAHEU

LOGIN: Céline21



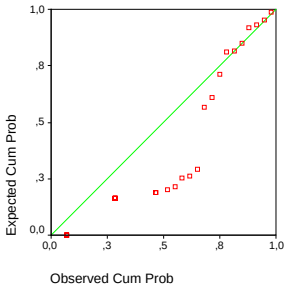
Gamma P-P Plot of DELTAHEU

LOGIN: chap's



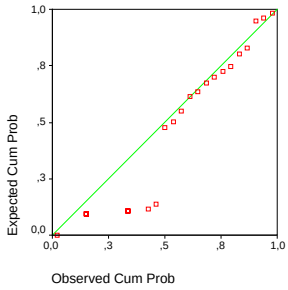
Gamma P-P Plot of DELTAHEU

LOGIN: Chrispi1



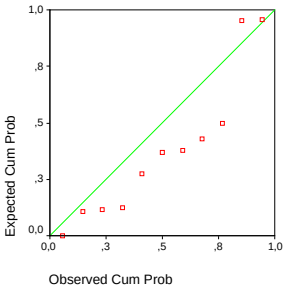
Gamma P-P Plot of DELTAHEU

LOGIN: clemisa



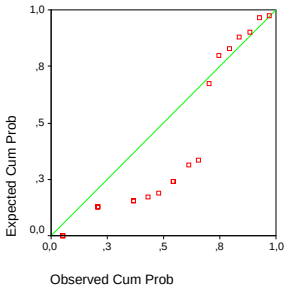
Gamma P-P Plot of DELTAHEU

LOGIN: Cypher



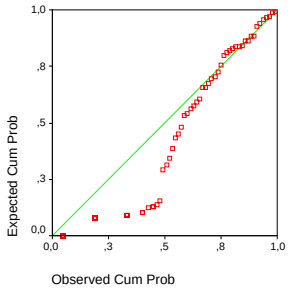
Gamma P-P Plot of DELTAHEU

LOGIN: DALMATE



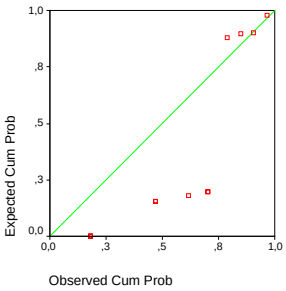
Gamma P-P Plot of DELTAHEU

LOGIN: Delatte



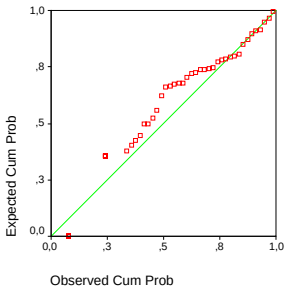
Gamma P-P Plot of DELTAHEU

LOGIN: didine77



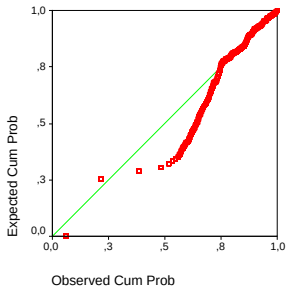
Gamma P-P Plot of DELTAHEU

LOGIN: Dinou



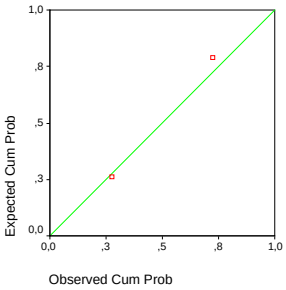
Gamma P-P Plot of DELTAHEU

LOGIN: djoule



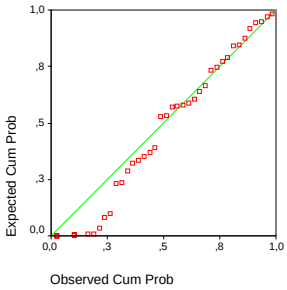
Gamma P-P Plot of DELTAHEU

LOGIN: dolores3



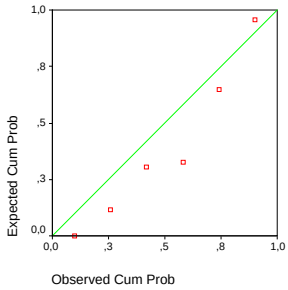
Gamma P-P Plot of DELTAHEU

LOGIN: dperrot



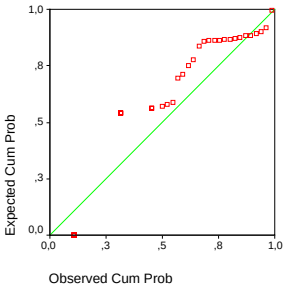
Gamma P-P Plot of DELTAHEU

LOGIN: dzhouari



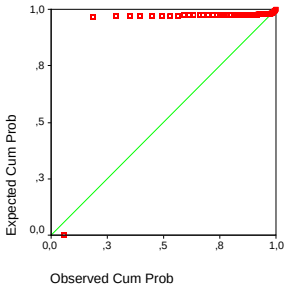
Gamma P-P Plot of DELTAHEU

LOGIN: ehran



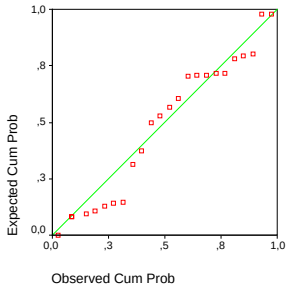
Gamma P-P Plot of DELTAHEU

LOGIN: Emilie



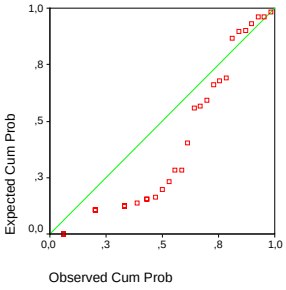
Gamma P-P Plot of DELTAHEU

LOGIN: eusebio1



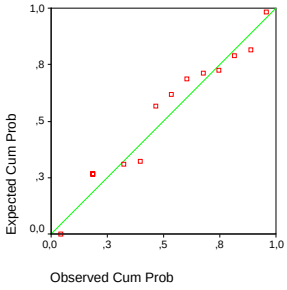
Gamma P-P Plot of DELTAHEU

LOGIN: fabby93



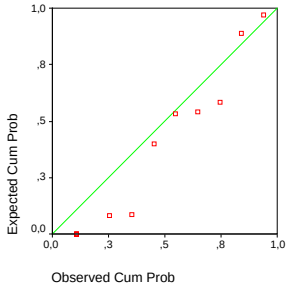
Gamma P-P Plot of DELTAHEU

LOGIN: fabou



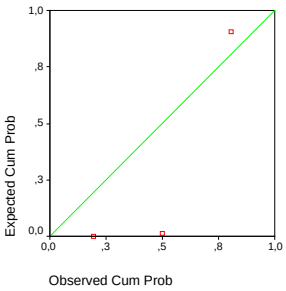
Gamma P-P Plot of DELTAHEU

LOGIN: fallouis



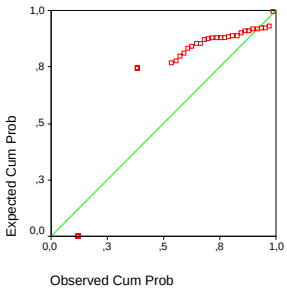
Gamma P-P Plot of DELTAHEU

LOGIN: fevrard



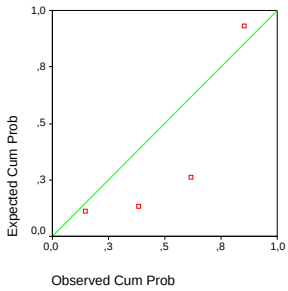
Gamma P-P Plot of DELTAHEU

LOGIN: FOURNIAL



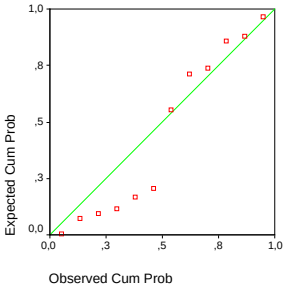
Gamma P-P Plot of DELTAHEU

LOGIN: Fruor



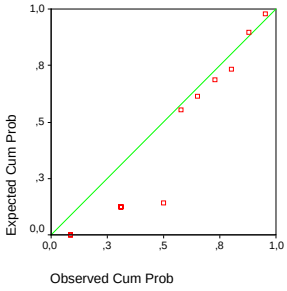
Gamma P-P Plot of DELTAHEU

LOGIN: frzpower



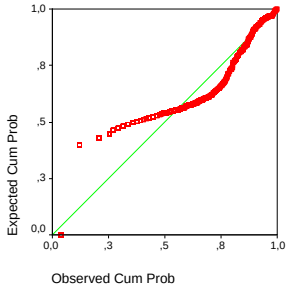
Gamma P-P Plot of DELTAHEU

LOGIN: GaëlleM



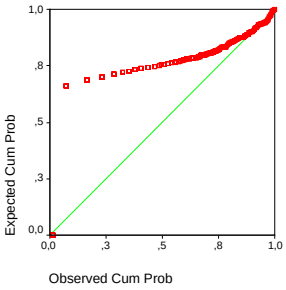
Gamma P-P Plot of DELTAHEU

LOGIN: gallouin



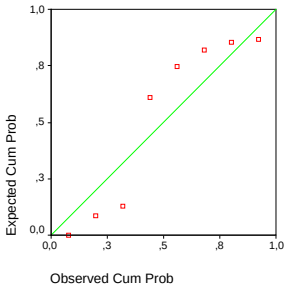
Gamma P-P Plot of DELTAHEU

LOGIN: Ginie31



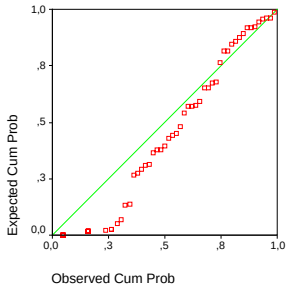
Gamma P-P Plot of DELTAHEU

LOGIN: gougay



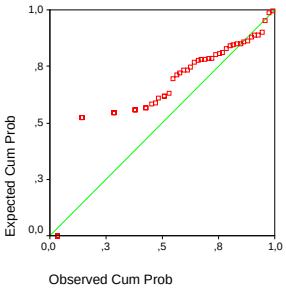
Gamma P-P Plot of DELTAHEU

LOGIN: Greensky



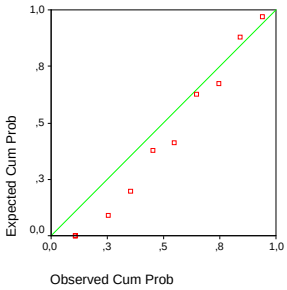
Gamma P-P Plot of DELTAHEU

LOGIN: gzing



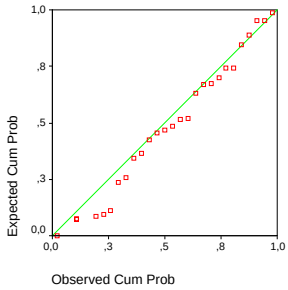
Gamma P-P Plot of DELTAHEU

LOGIN: hedi.dc



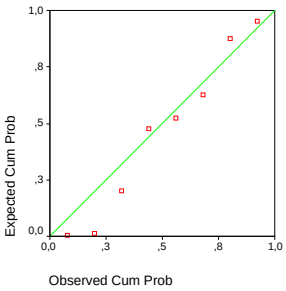
Gamma P-P Plot of DELTAHEU

LOGIN: hutch



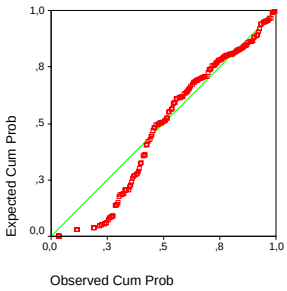
Gamma P-P Plot of DELTAHEU

LOGIN: jeanbask



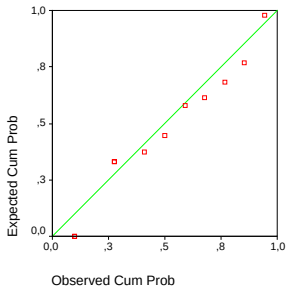
Gamma P-P Plot of DELTAHEU

LOGIN: jeannot3



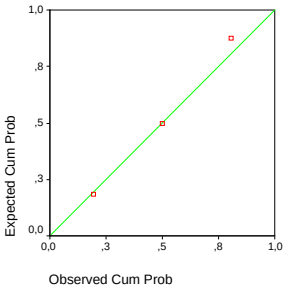
Gamma P-P Plot of DELTAHEU

LOGIN: joewj



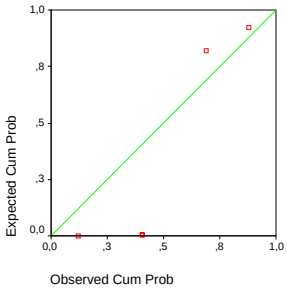
Gamma P-P Plot of DELTAHEU

LOGIN: juboubou



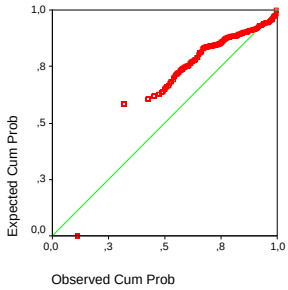
Gamma P-P Plot of DELTAHEU

LOGIN: juju75



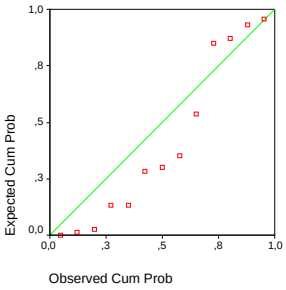
Gamma P-P Plot of DELTAHEU

LOGIN: kabotine



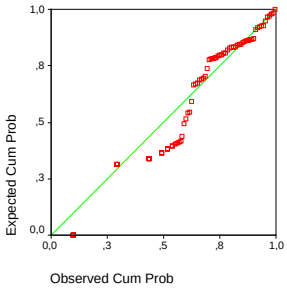
Gamma P-P Plot of DELTAHEU

LOGIN: Kathrin



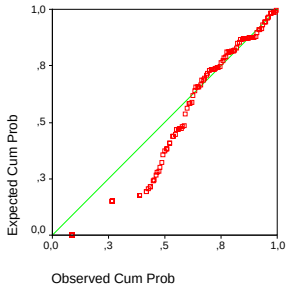
Gamma P-P Plot of DELTAHEU

LOGIN: Kifran



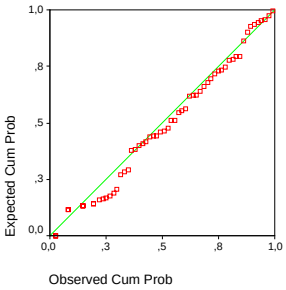
Gamma P-P Plot of DELTAHEU

LOGIN: lbarbier



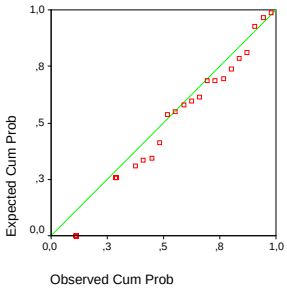
Gamma P-P Plot of DELTAHEU

LOGIN: levieux



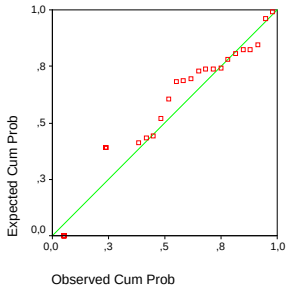
Gamma P-P Plot of DELTAHEU

LOGIN: LoïcT



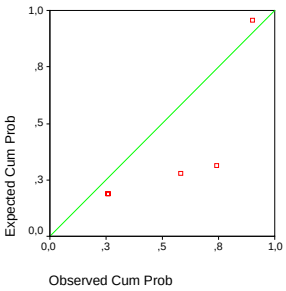
Gamma P-P Plot of DELTAHEU

LOGIN: Loquet



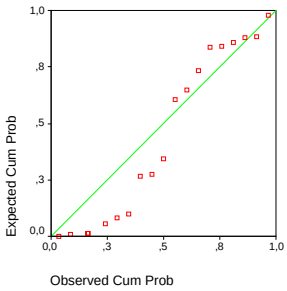
Gamma P-P Plot of DELTAHEU

LOGIN: luberdov



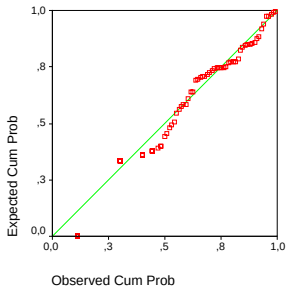
Gamma P-P Plot of DELTAHEU

LOGIN: ludovicb



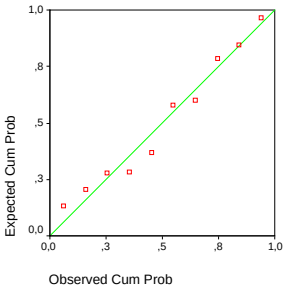
Gamma P-P Plot of DELTAHEU

LOGIN: m@hy



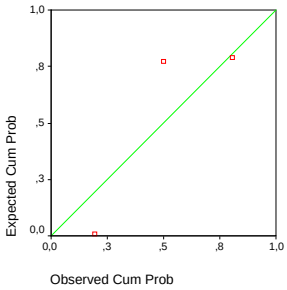
Gamma P-P Plot of DELTAHEU

LOGIN: manwe



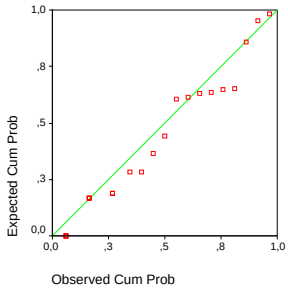
Gamma P-P Plot of DELTAHEU

LOGIN: mary_ce



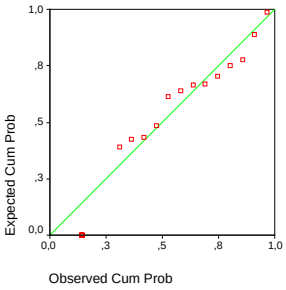
Gamma P-P Plot of DELTAHEU

LOGIN: Mayfair



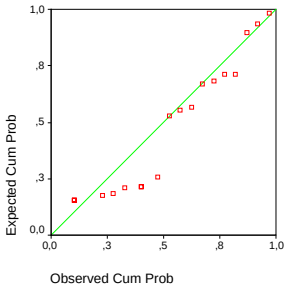
Gamma P-P Plot of DELTAHEU

LOGIN: mekiela



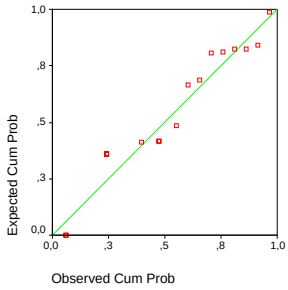
Gamma P-P Plot of DELTAHEU

LOGIN: Michel



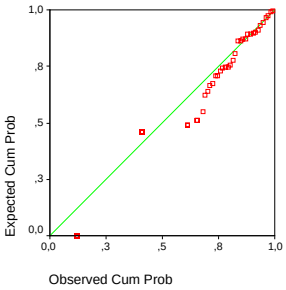
Gamma P-P Plot of DELTAHEU

LOGIN: micka74



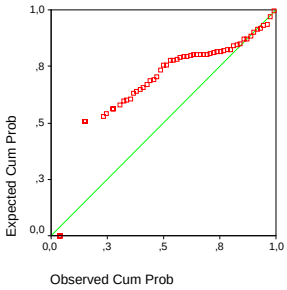
Gamma P-P Plot of DELTAHEU

LOGIN: mike_s



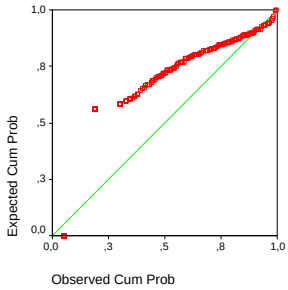
Gamma P-P Plot of DELTAHEU

LOGIN: minious



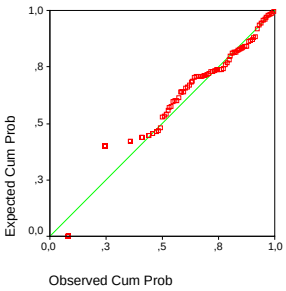
Gamma P-P Plot of DELTAHEU

LOGIN: mperrot



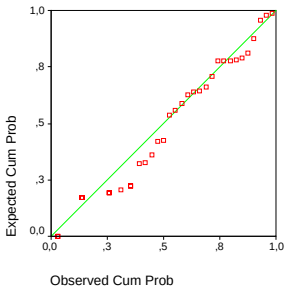
Gamma P-P Plot of DELTAHEU

LOGIN: Munier



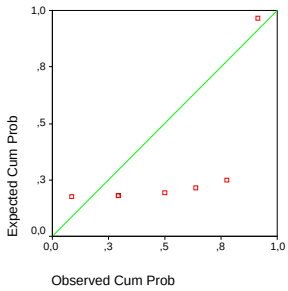
Gamma P-P Plot of DELTAHEU

LOGIN: Naemesis



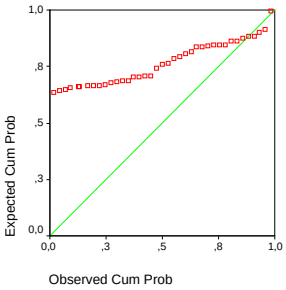
Gamma P-P Plot of DELTAHEU

LOGIN: nhai07



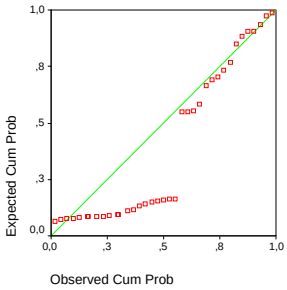
Gamma P-P Plot of DELTAHEU

LOGIN: nickus



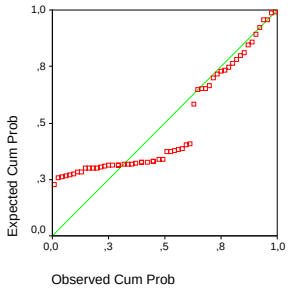
Gamma P-P Plot of DELTAHEU

LOGIN: noel_net



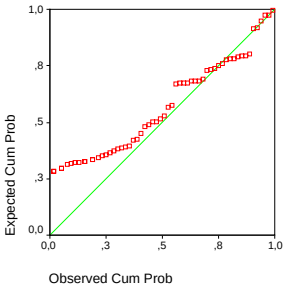
Gamma P-P Plot of DELTAHEU

LOGIN: philia



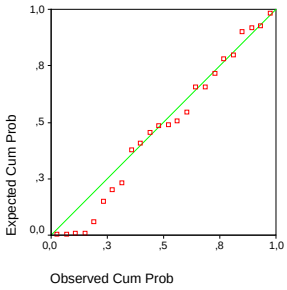
Gamma P-P Plot of DELTAHEU

LOGIN: popcorn



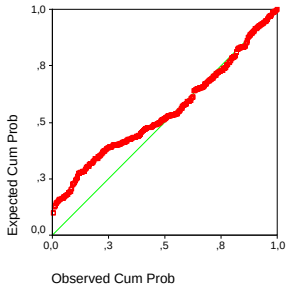
Gamma P-P Plot of DELTAHEU

LOGIN: prioton



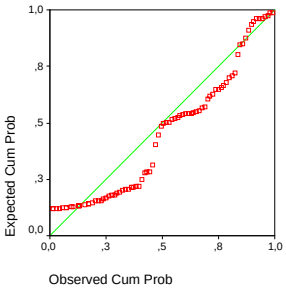
Gamma P-P Plot of DELTAHEU

LOGIN: prozaiik



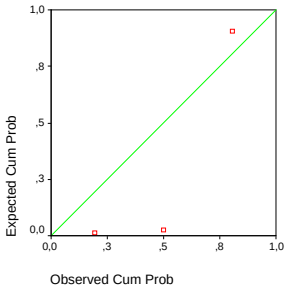
Gamma P-P Plot of DELTAHEU

LOGIN: quentin



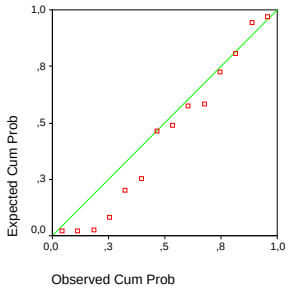
Gamma P-P Plot of DELTAHEU

LOGIN: RAPH



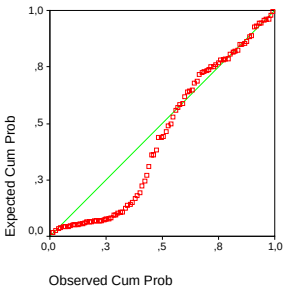
Gamma P-P Plot of DELTAHEU

LOGIN: relindis



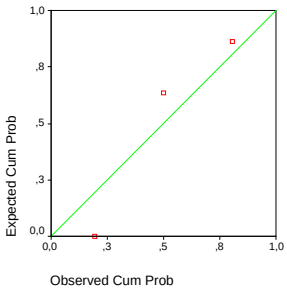
Gamma P-P Plot of DELTAHEU

LOGIN: rlaval



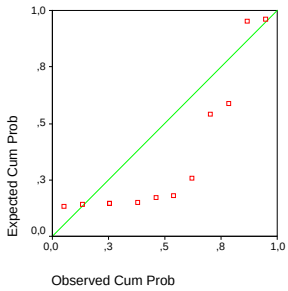
Gamma P-P Plot of DELTAHEU

LOGIN: sandor



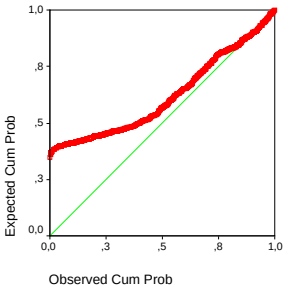
Gamma P-P Plot of DELTAHEU

LOGIN: sebulba9



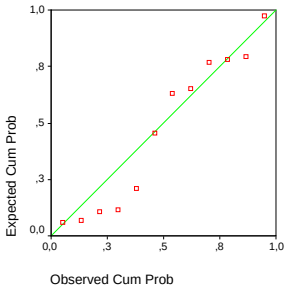
Gamma P-P Plot of DELTAHEU

LOGIN: semery



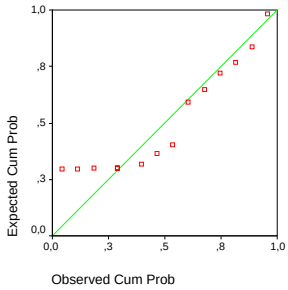
Gamma P-P Plot of DELTAHEU

LOGIN: sicart



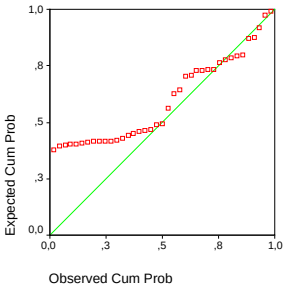
Gamma P-P Plot of DELTAHEU

LOGIN: sophynet



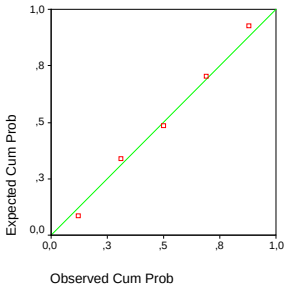
Gamma P-P Plot of DELTAHEU

LOGIN: Stèph



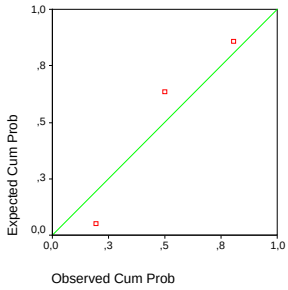
Gamma P-P Plot of DELTAHEU

LOGIN: stumpf



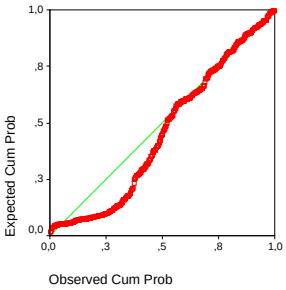
Gamma P-P Plot of DELTAHEU

LOGIN: Sylvie4



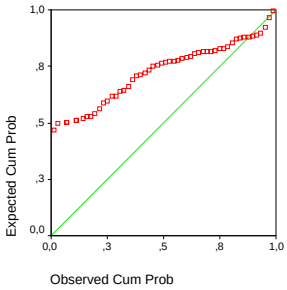
Gamma P-P Plot of DELTAHEU

LOGIN: Thibaut



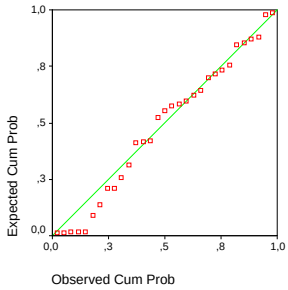
Gamma P-P Plot of DELTAHEU

LOGIN: Tonio2k



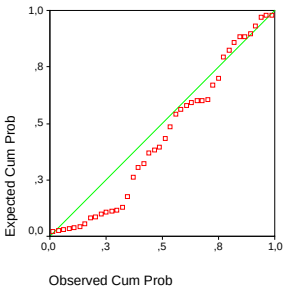
Gamma P-P Plot of DELTAHEU

LOGIN: toto2



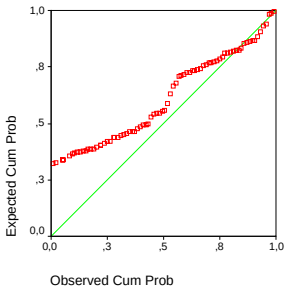
Gamma P-P Plot of DELTAHEU

LOGIN: tranb



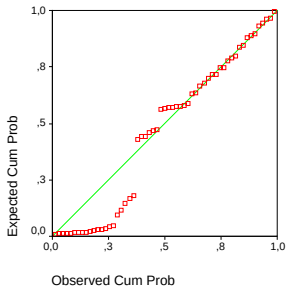
Gamma P-P Plot of DELTAHEU

LOGIN: vianney2



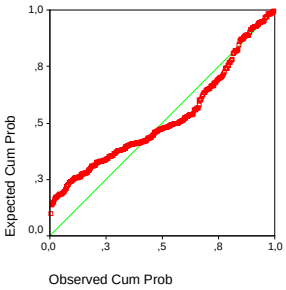
Gamma P-P Plot of DELTAHEU

LOGIN: Virginie



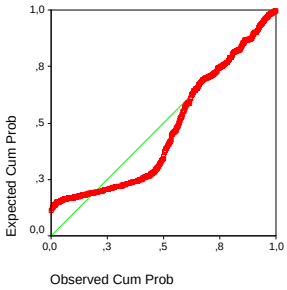
Gamma P-P Plot of DELTAHEU

LOGIN: xzouf



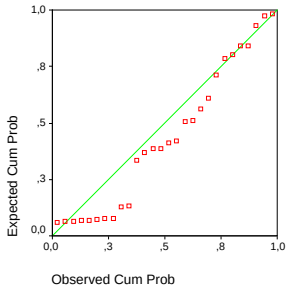
Gamma P-P Plot of DELTAHEU

LOGIN: Yseulys



Gamma P-P Plot of DELTAHEU

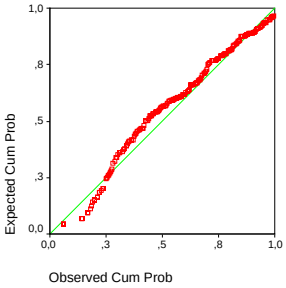
LOGIN: Yves



PP PLOT LOI LOG-NORMALE:

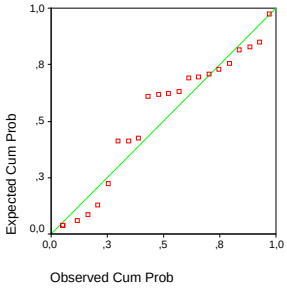
Lognormal P-P Plot of DELTAHI

LOGIN: 130961



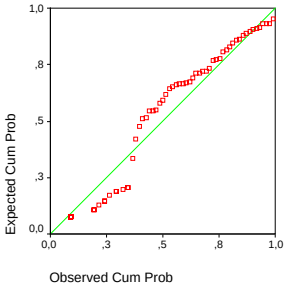
Lognormal P-P Plot of DELTAHI

LOGIN: adribou



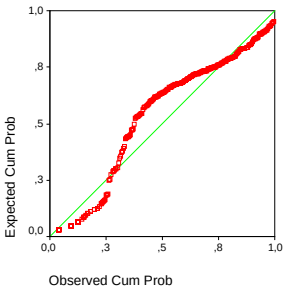
Lognormal P-P Plot of DELTAHI

LOGIN: akwell1



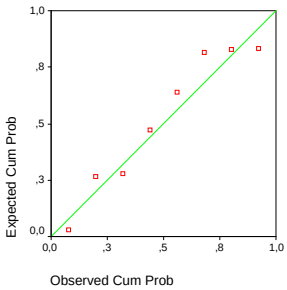
Lognormal P-P Plot of DELTAHI

LOGIN: AlexO



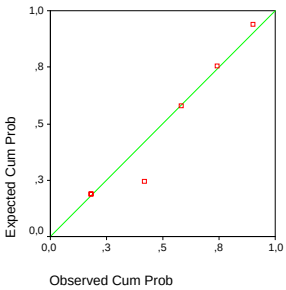
Lognormal P-P Plot of DELTAHI

LOGIN: ama1



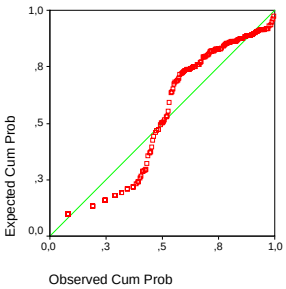
Lognormal P-P Plot of DELTAHI

LOGIN: amibio



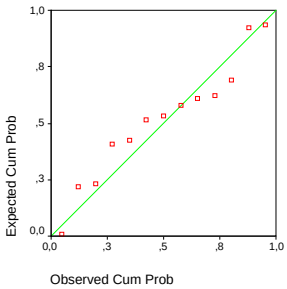
Lognormal P-P Plot of DELTAHI

LOGIN: AnnA



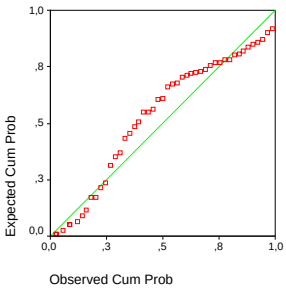
Lognormal P-P Plot of DELTAHI

LOGIN: apopo



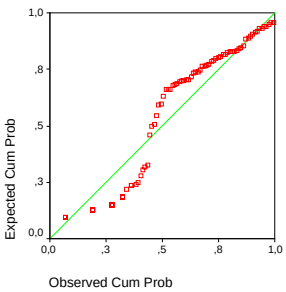
Lognormal P-P Plot of DELTAHI

LOGIN: ashley85



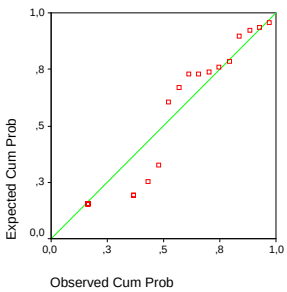
Lognormal P-P Plot of DELTAHI

LOGIN: Aude



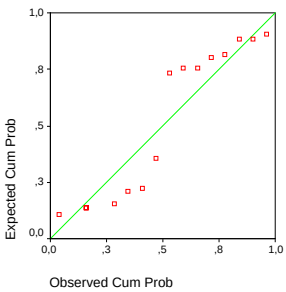
Lognormal P-P Plot of DELTAHI

LOGIN: badir



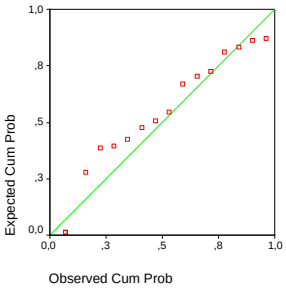
Lognormal P-P Plot of DELTAHI

LOGIN: bgougay



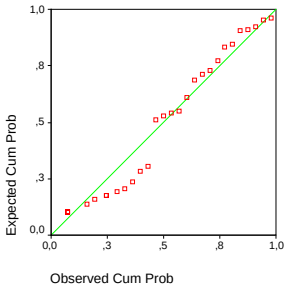
Lognormal P-P Plot of DELTAHI

LOGIN: blondi



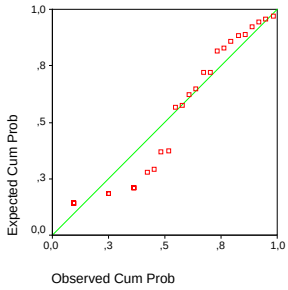
Lognormal P-P Plot of DELTAHI

LOGIN: bontze



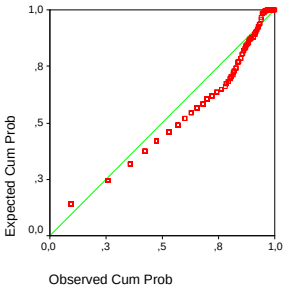
Lognormal P-P Plot of DELTAHI

LOGIN: bpdl



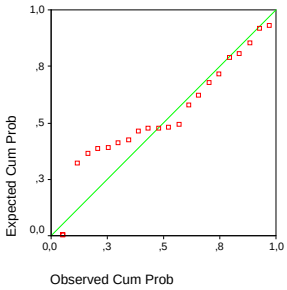
Lognormal P-P Plot of DELTAHI

LOGIN: brusson1



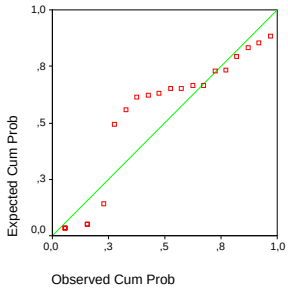
Lognormal P-P Plot of DELTAHI

LOGIN: bruyere



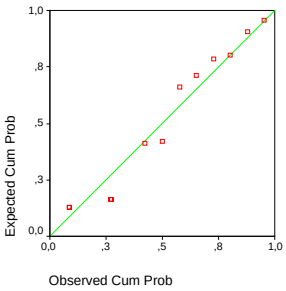
Lognormal P-P Plot of DELTAHI

LOGIN: carolala



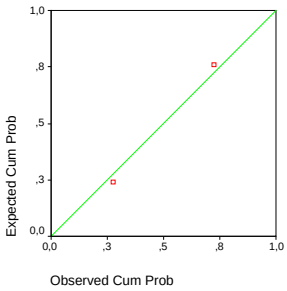
Lognormal P-P Plot of DELTAHI

LOGIN: cathou1



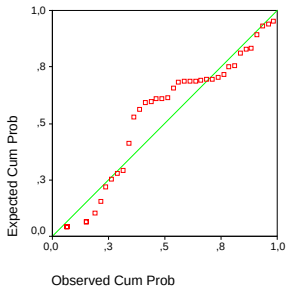
Lognormal P-P Plot of DELTAHI

LOGIN: CATHOU25



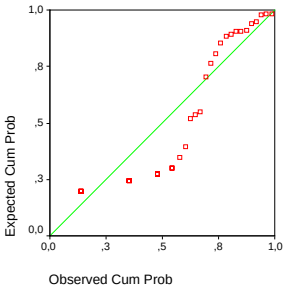
Lognormal P-P Plot of DELTAHI

LOGIN: Céline21



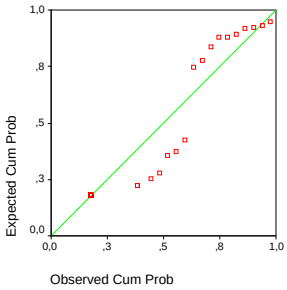
Lognormal P-P Plot of DELTAHI

LOGIN: chap's



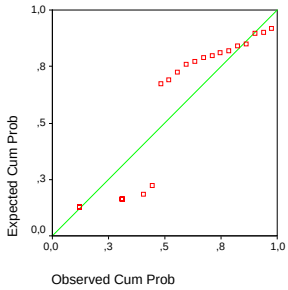
Lognormal P-P Plot of DELTAHI

LOGIN: Chrispi1



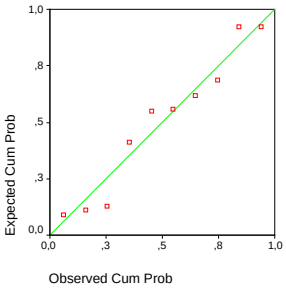
Lognormal P-P Plot of DELTAHI

LOGIN: clemisa



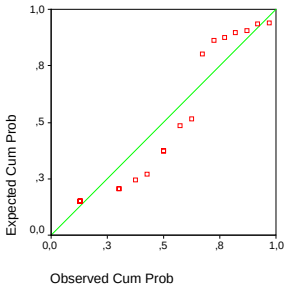
Lognormal P-P Plot of DELTAHI

LOGIN: Cypher



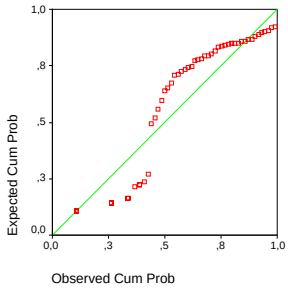
Lognormal P-P Plot of DELTAHI

LOGIN: DALMATE



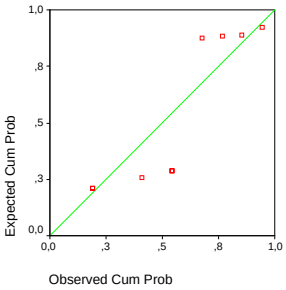
Lognormal P-P Plot of DELTAHI

LOGIN: Delatte



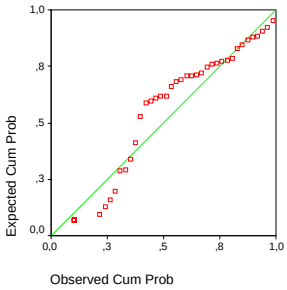
Lognormal P-P Plot of DELTAHI

LOGIN: didine77



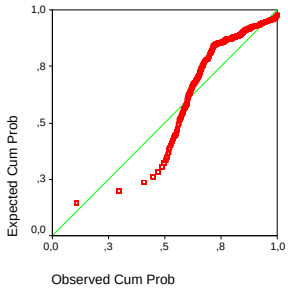
Lognormal P-P Plot of DELTAHI

LOGIN: Dinou



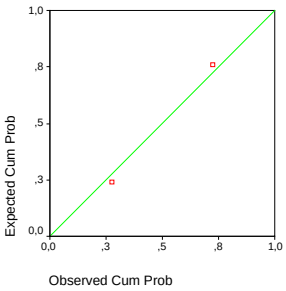
Lognormal P-P Plot of DELTAHI

LOGIN: djoule



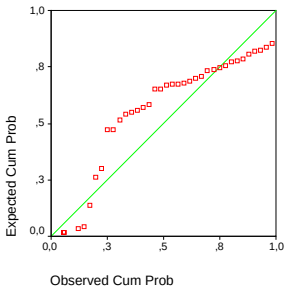
Lognormal P-P Plot of DELTAHI

LOGIN: dolores3



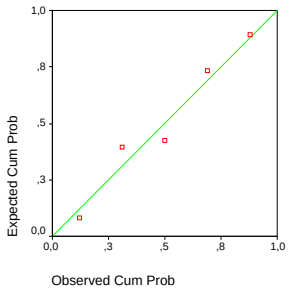
Lognormal P-P Plot of DELTAHI

LOGIN: dperrot



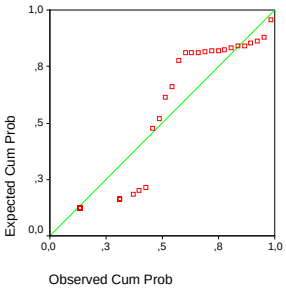
Lognormal P-P Plot of DELTAHI

LOGIN: dzhouari



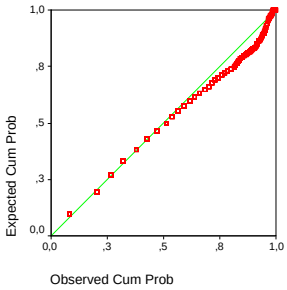
Lognormal P-P Plot of DELTAHI

LOGIN: ehran



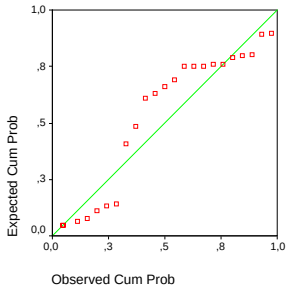
Lognormal P-P Plot of DELTAHI

LOGIN: Emilie



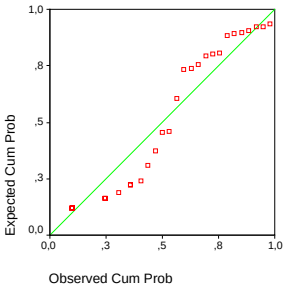
Lognormal P-P Plot of DELTAHI

LOGIN: eusebio1



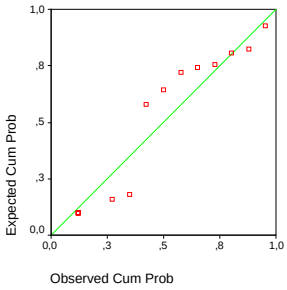
Lognormal P-P Plot of DELTAHI

LOGIN: fabby93



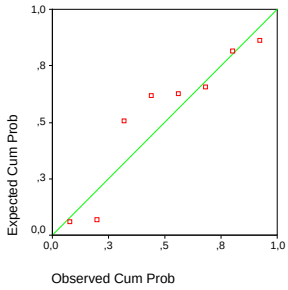
Lognormal P-P Plot of DELTAHI

LOGIN: fabou



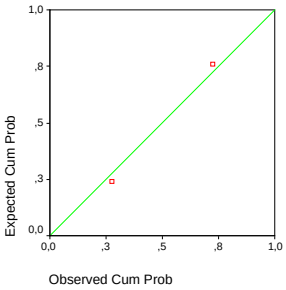
Lognormal P-P Plot of DELTAHI

LOGIN: fallouis



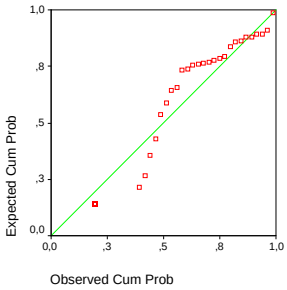
Lognormal P-P Plot of DELTAHI

LOGIN: fevrard



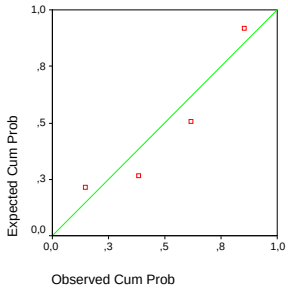
Lognormal P-P Plot of DELTAHI

LOGIN: FOURNIAL



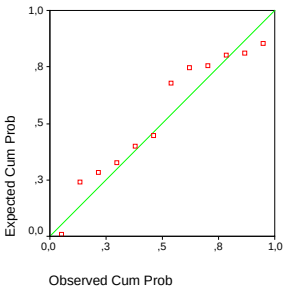
Lognormal P-P Plot of DELTAHI

LOGIN: Fruor



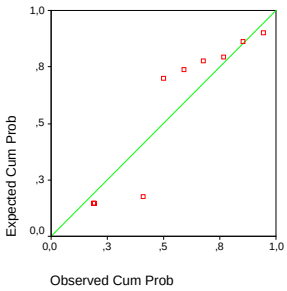
Lognormal P-P Plot of DELTAHI

LOGIN: frzpower



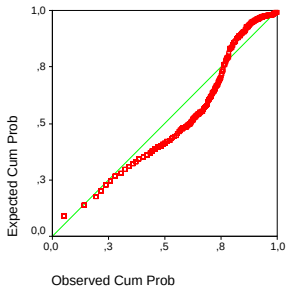
Lognormal P-P Plot of DELTAHI

LOGIN: GaëlleM



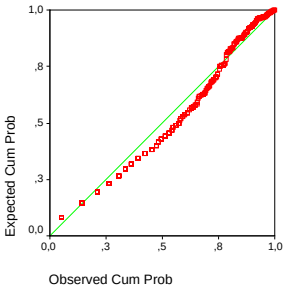
Lognormal P-P Plot of DELTAHI

LOGIN: gallouin



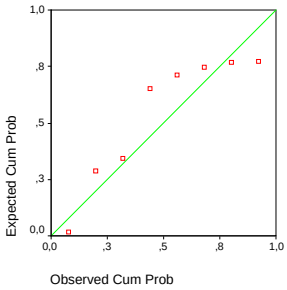
Lognormal P-P Plot of DELTAHI

LOGIN: Ginie31



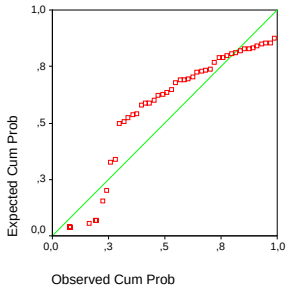
Lognormal P-P Plot of DELTAHI

LOGIN: gougay



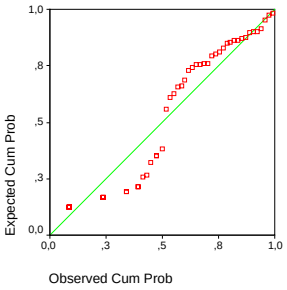
Lognormal P-P Plot of DELTAHI

LOGIN: Greensky



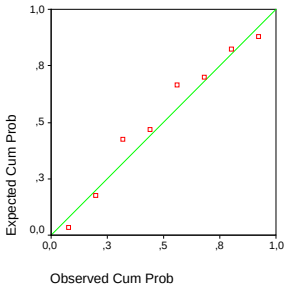
Lognormal P-P Plot of DELTAHI

LOGIN: gzing



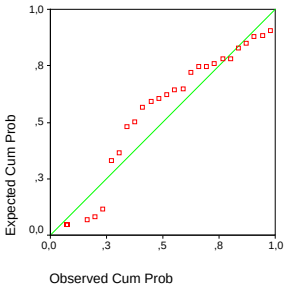
Lognormal P-P Plot of DELTAHI

LOGIN: hedi.dc



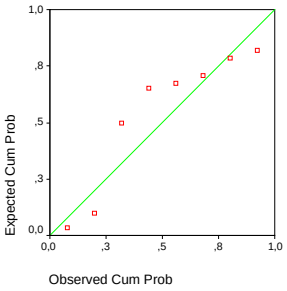
Lognormal P-P Plot of DELTAHI

LOGIN: hutch



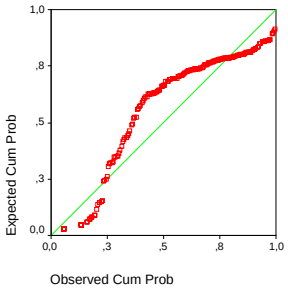
Lognormal P-P Plot of DELTAHI

LOGIN: jeanbask



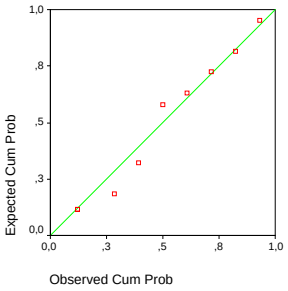
Lognormal P-P Plot of DELTAHI

LOGIN: jeannot3



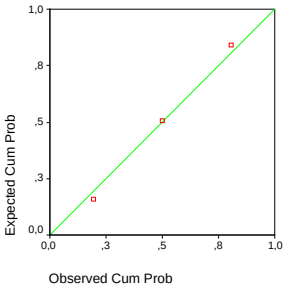
Lognormal P-P Plot of DELTAHI

LOGIN: joewj



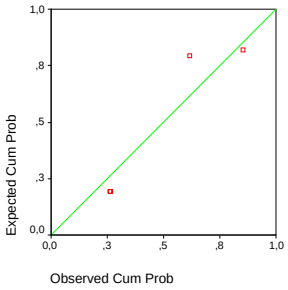
Lognormal P-P Plot of DELTAHI

LOGIN: juboubou



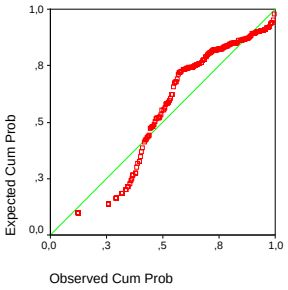
Lognormal P-P Plot of DELTAHI

LOGIN: juju75



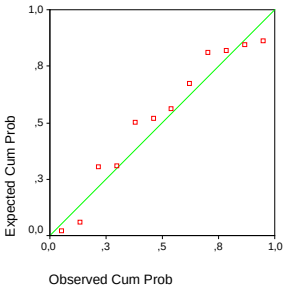
Lognormal P-P Plot of DELTAHI

LOGIN: kabotine



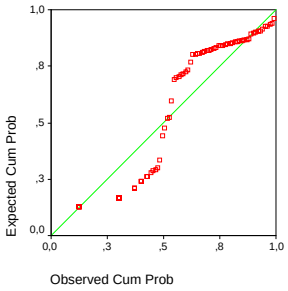
Lognormal P-P Plot of DELTAHI

LOGIN: Kathrin



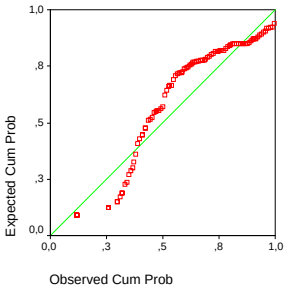
Lognormal P-P Plot of DELTAHI

LOGIN: Kifran



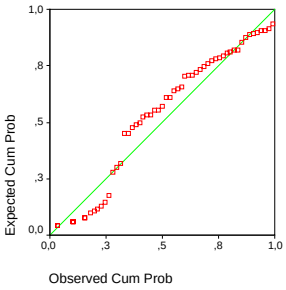
Lognormal P-P Plot of DELTAHI

LOGIN: lbarbier



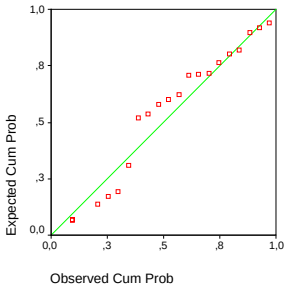
Lognormal P-P Plot of DELTAHI

LOGIN: levieux



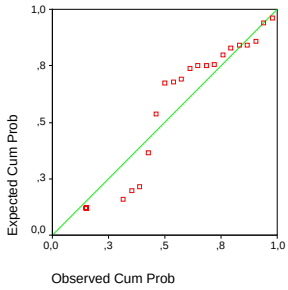
Lognormal P-P Plot of DELTAHI

LOGIN: LoïcT



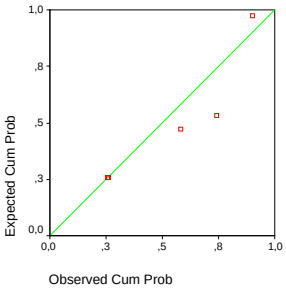
Lognormal P-P Plot of DELTAHI

LOGIN: Loquet



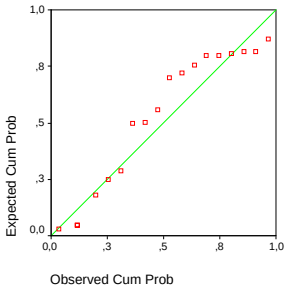
Lognormal P-P Plot of DELTAHI

LOGIN: luberdov



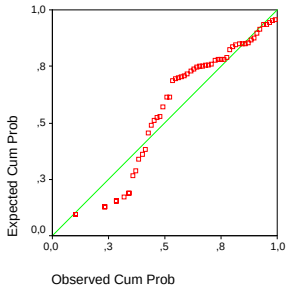
Lognormal P-P Plot of DELTAHI

LOGIN: ludovicb



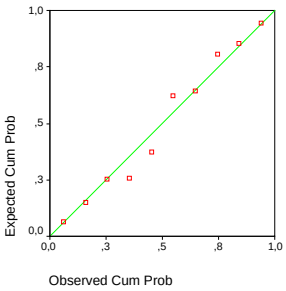
Lognormal P-P Plot of DELTAHI

LOGIN: m@hy



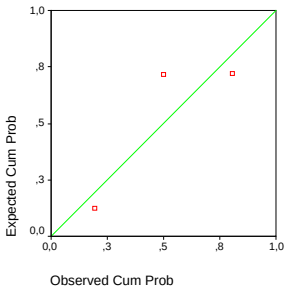
Lognormal P-P Plot of DELTAHI

LOGIN: manwe



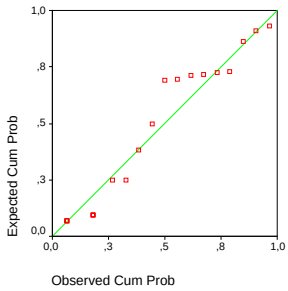
Lognormal P-P Plot of DELTAHI

LOGIN: mary_ce



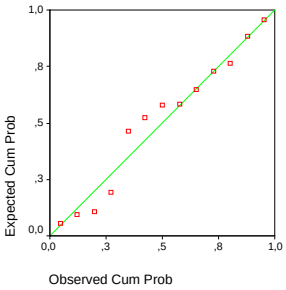
Lognormal P-P Plot of DELTAHI

LOGIN: Mayfair



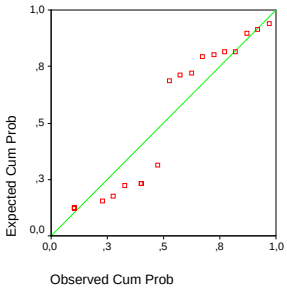
Lognormal P-P Plot of DELTAHI

LOGIN: mekiela



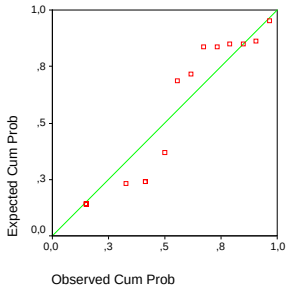
Lognormal P-P Plot of DELTAHI

LOGIN: Michel



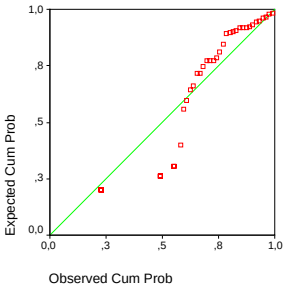
Lognormal P-P Plot of DELTAHI

LOGIN: micka74



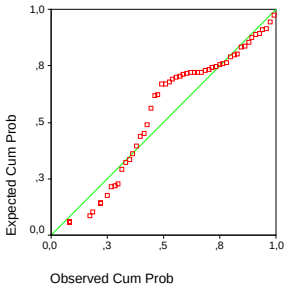
Lognormal P-P Plot of DELTAHI

LOGIN: mike_s



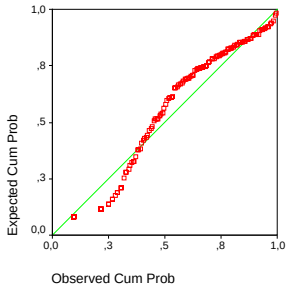
Lognormal P-P Plot of DELTAHI

LOGIN: minious



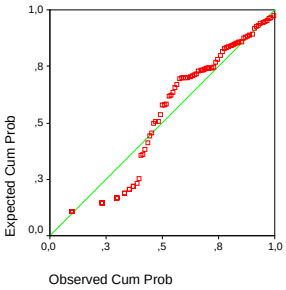
Lognormal P-P Plot of DELTAHI

LOGIN: mperrot



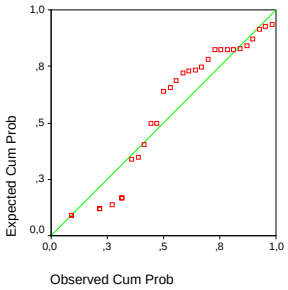
Lognormal P-P Plot of DELTAHI

LOGIN: Munier



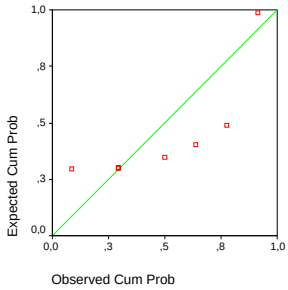
Lognormal P-P Plot of DELTAHI

LOGIN: Naemesis



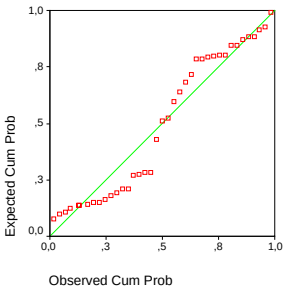
Lognormal P-P Plot of DELTAHI

LOGIN: nhai07



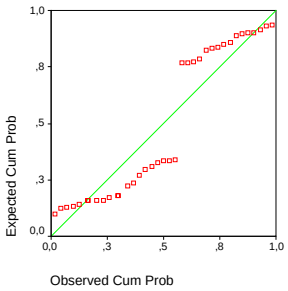
Lognormal P-P Plot of DELTAHI

LOGIN: nickus



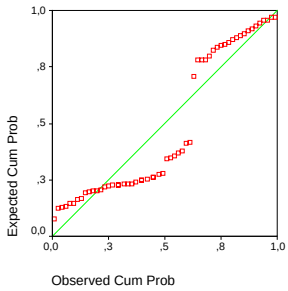
Lognormal P-P Plot of DELTAHI

LOGIN: noel_net



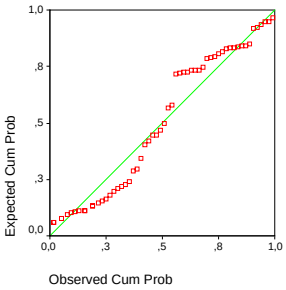
Lognormal P-P Plot of DELTAHI

LOGIN: philia



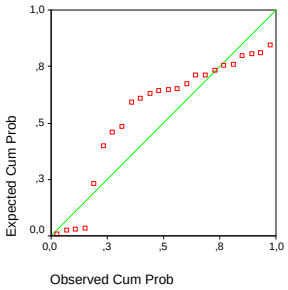
Lognormal P-P Plot of DELTAHI

LOGIN: popcorn



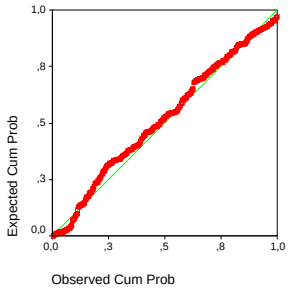
Lognormal P-P Plot of DELTAHI

LOGIN: prioton



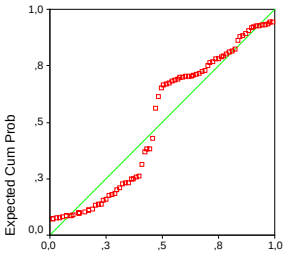
Lognormal P-P Plot of DELTAHI

LOGIN: prozaiik



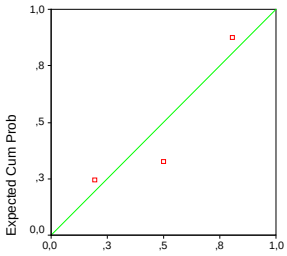
Lognormal P-P Plot of DELTAHI

LOGIN: quentin



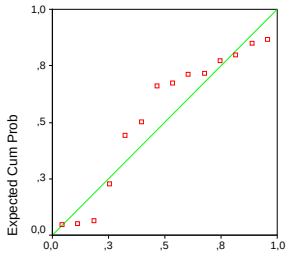
Lognormal P-P Plot of DELTAHI

LOGIN: RAPH



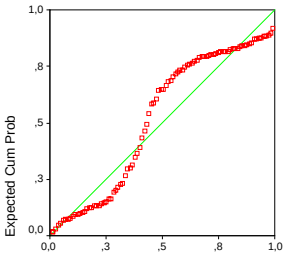
Lognormal P-P Plot of DELTAHI

LOGIN: relindis



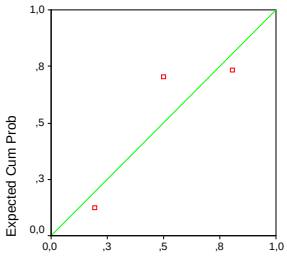
Lognormal P-P Plot of DELTAHI

LOGIN: rlaval



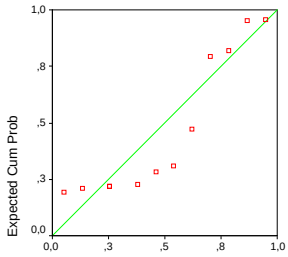
Lognormal P-P Plot of DELTAHI

LOGIN: sandor



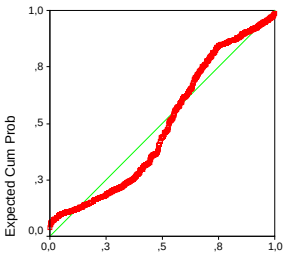
Lognormal P-P Plot of DELTAHI

LOGIN: sebulba9



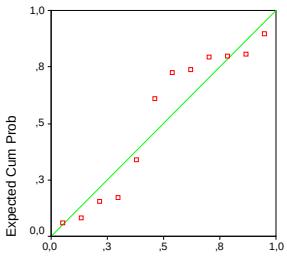
Lognormal P-P Plot of DELTAHI

LOGIN: semery



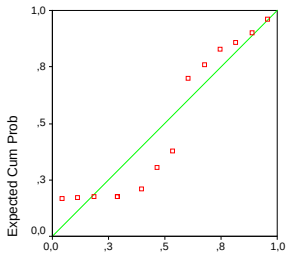
Lognormal P-P Plot of DELTAHI

LOGIN: sicart



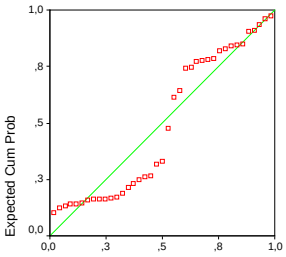
Lognormal P-P Plot of DELTAHI

LOGIN: sophynet



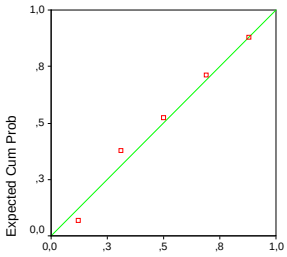
Lognormal P-P Plot of DELTAHI

LOGIN: Stèph



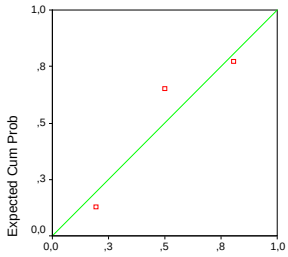
Lognormal P-P Plot of DELTAHI

LOGIN: stumpf



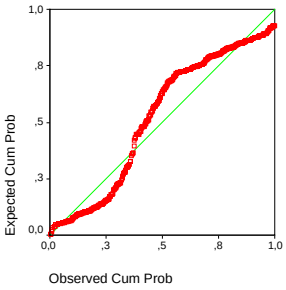
Lognormal P-P Plot of DELTAHI

LOGIN: Sylvie4



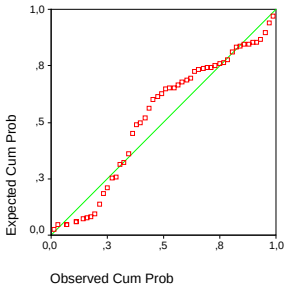
Lognormal P-P Plot of DELTAHI

LOGIN: Thibaut



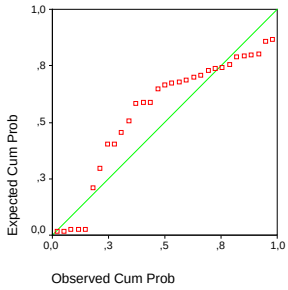
Lognormal P-P Plot of DELTAHI

LOGIN: Tonio2k



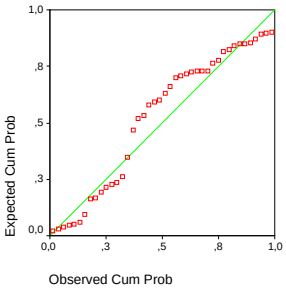
Lognormal P-P Plot of DELTAHI

LOGIN: toto2



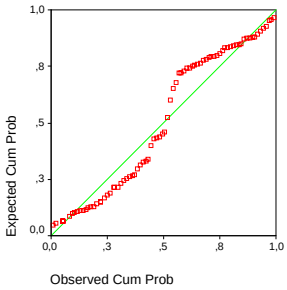
Lognormal P-P Plot of DELTAHI

LOGIN: tranb



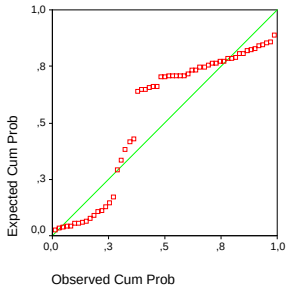
Lognormal P-P Plot of DELTAHI

LOGIN: vianney2



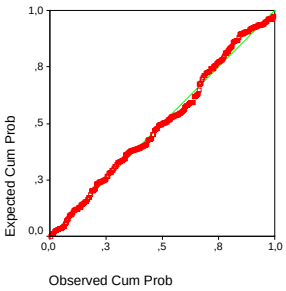
Lognormal P-P Plot of DELTAHI

LOGIN: Virginie



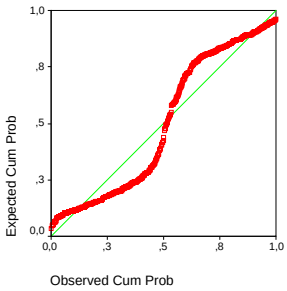
Lognormal P-P Plot of DELTAHI

LOGIN: xzouf



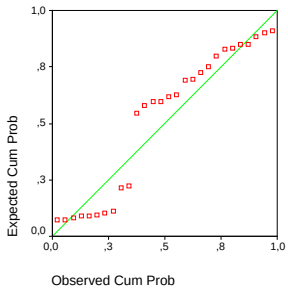
Lognormal P-P Plot of DELTAHI

LOGIN: Yseulys



Lognormal P-P Plot of DELTAHI

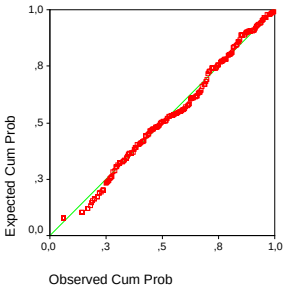
LOGIN: Yves



PP PLOT LOI WEIBULL:

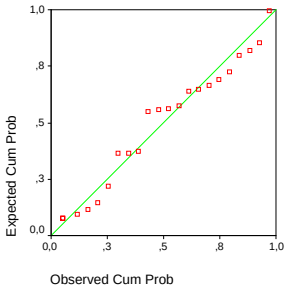
Weibull P-P Plot of DELTAHEU

LOGIN: 130961



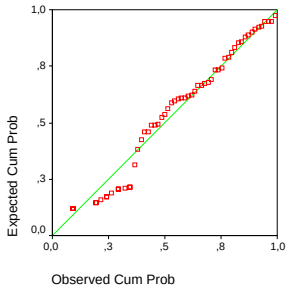
Weibull P-P Plot of DELTAHEU

LOGIN: adribou



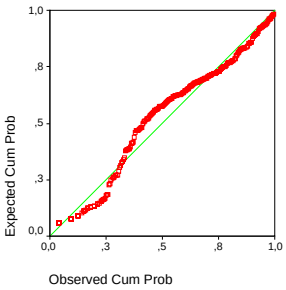
Weibull P-P Plot of DELTAHEU

LOGIN: akwell1



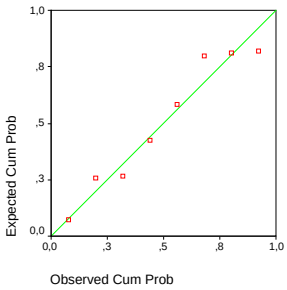
Weibull P-P Plot of DELTAHEU

LOGIN: AlexO



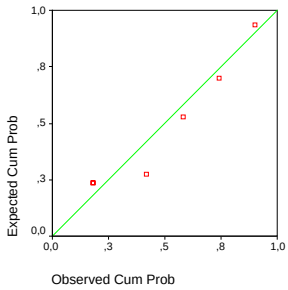
Weibull P-P Plot of DELTAHEU

LOGIN: ama1



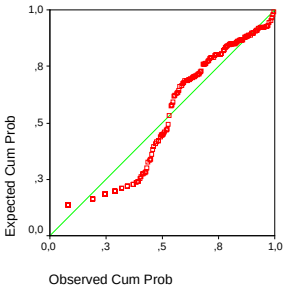
Weibull P-P Plot of DELTAHEU

LOGIN: amibio



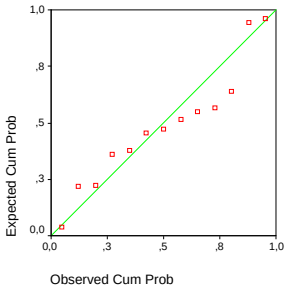
Weibull P-P Plot of DELTAHEU

LOGIN: AnnA



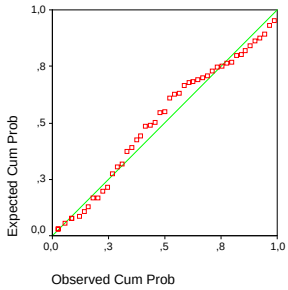
Weibull P-P Plot of DELTAHEU

LOGIN: apopo



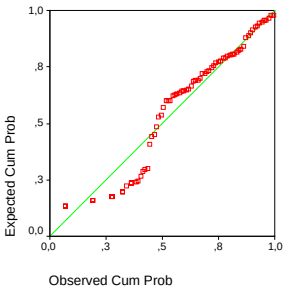
Weibull P-P Plot of DELTAHEU

LOGIN: ashley85



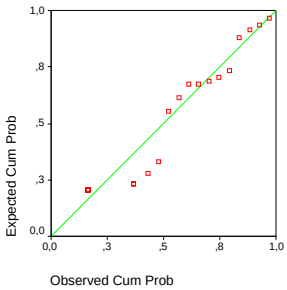
Weibull P-P Plot of DELTAHEU

LOGIN: Aude



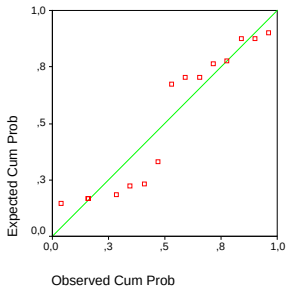
Weibull P-P Plot of DELTAHEU

LOGIN: badir



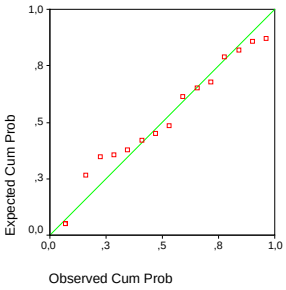
Weibull P-P Plot of DELTAHEU

LOGIN: bgougay



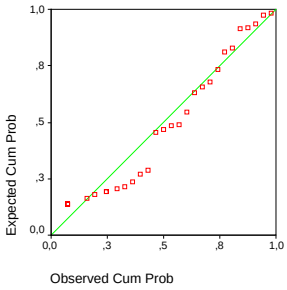
Weibull P-P Plot of DELTAHEU

LOGIN: blondi



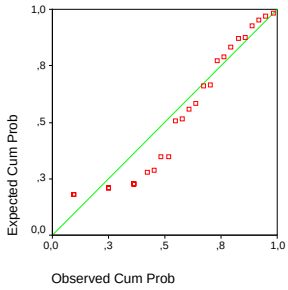
Weibull P-P Plot of DELTAHEU

LOGIN: bontze



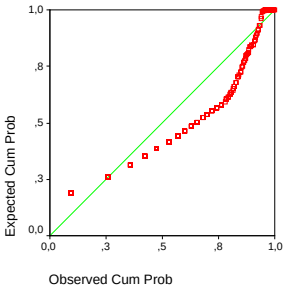
Weibull P-P Plot of DELTAHEU

LOGIN: bpdl



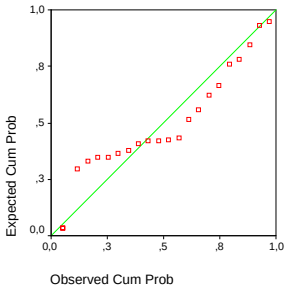
Weibull P-P Plot of DELTAHEU

LOGIN: brusson1



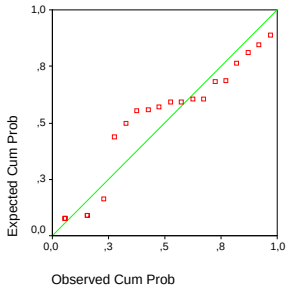
Weibull P-P Plot of DELTAHEU

LOGIN: bruyere



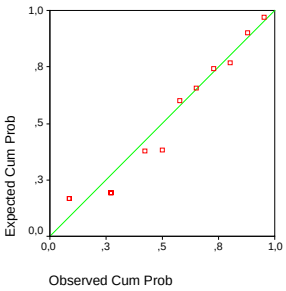
Weibull P-P Plot of DELTAHEU

LOGIN: carolala



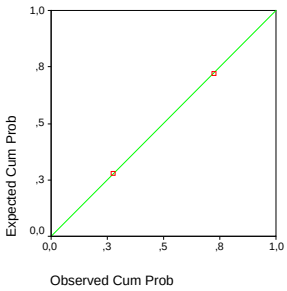
Weibull P-P Plot of DELTAHEU

LOGIN: cathou1



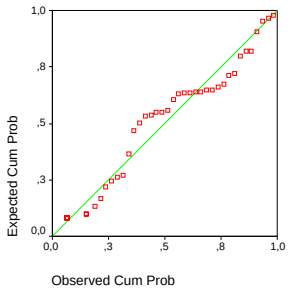
Weibull P-P Plot of DELTAHEU

LOGIN: CATHOU25



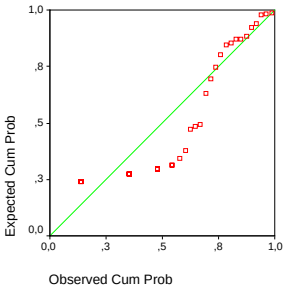
Weibull P-P Plot of DELTAHEU

LOGIN: Céline21



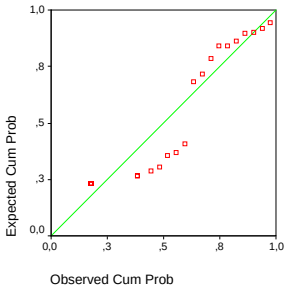
Weibull P-P Plot of DELTAHEU

LOGIN: chap's



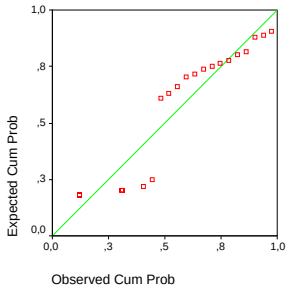
Weibull P-P Plot of DELTAHEU

LOGIN: Chrispi1



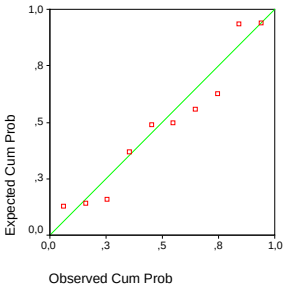
Weibull P-P Plot of DELTAHEU

LOGIN: clemisa



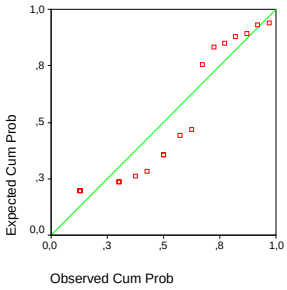
Weibull P-P Plot of DELTAHEU

LOGIN: Cypher



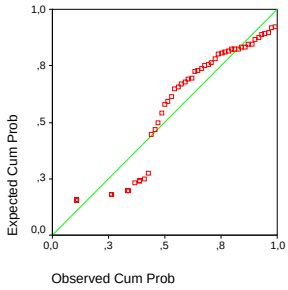
Weibull P-P Plot of DELTAHEU

LOGIN: DALMATE



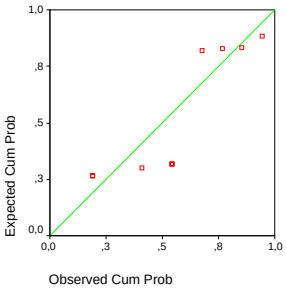
Weibull P-P Plot of DELTAHEU

LOGIN: Delatte



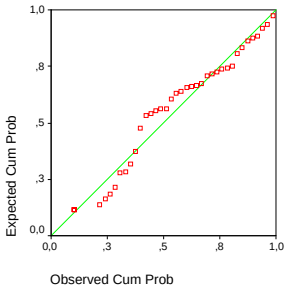
Weibull P-P Plot of DELTAHEU

LOGIN: didine77



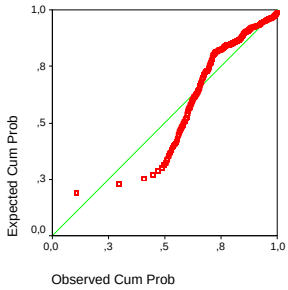
Weibull P-P Plot of DELTAHEU

LOGIN: Dinou



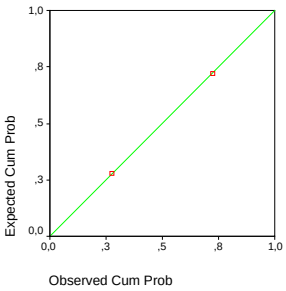
Weibull P-P Plot of DELTAHEU

LOGIN: djoule



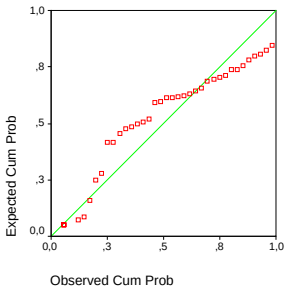
Weibull P-P Plot of DELTAHEU

LOGIN: dolores3



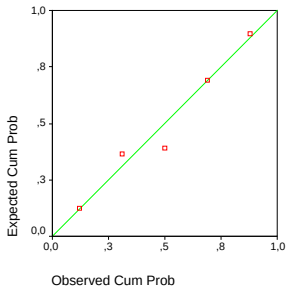
Weibull P-P Plot of DELTAHEU

LOGIN: dperrot



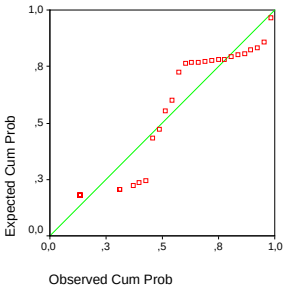
Weibull P-P Plot of DELTAHEU

LOGIN: dzhouari



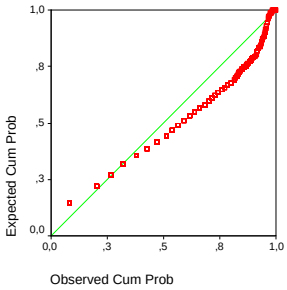
Weibull P-P Plot of DELTAHEU

LOGIN: ehran



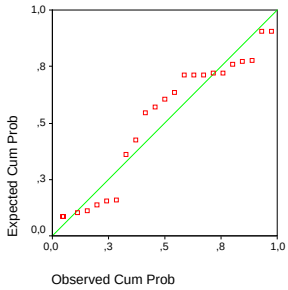
Weibull P-P Plot of DELTAHEU

LOGIN: Emilie



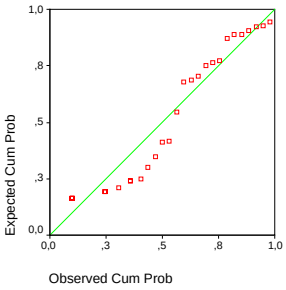
Weibull P-P Plot of DELTAHEU

LOGIN: eusebio1



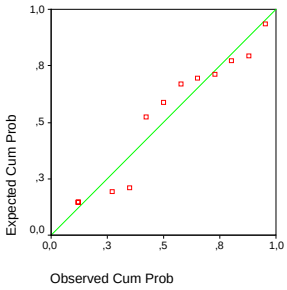
Weibull P-P Plot of DELTAHEU

LOGIN: fabby93



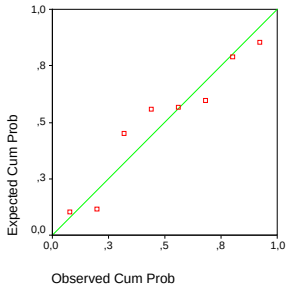
Weibull P-P Plot of DELTAHEU

LOGIN: fabou



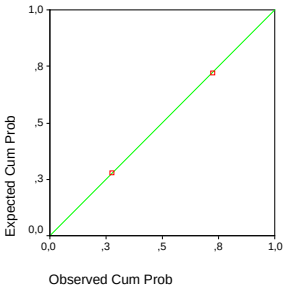
Weibull P-P Plot of DELTAHEU

LOGIN: fallouis



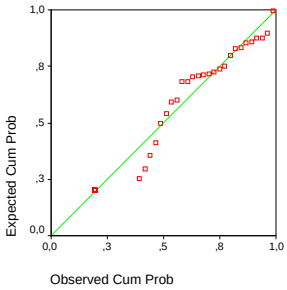
Weibull P-P Plot of DELTAHEU

LOGIN: fevrard



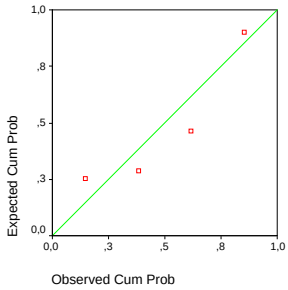
Weibull P-P Plot of DELTAHEU

LOGIN: FOURNIAL



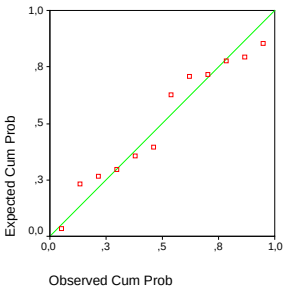
Weibull P-P Plot of DELTAHEU

LOGIN: Fruor



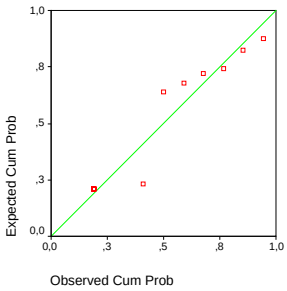
Weibull P-P Plot of DELTAHEU

LOGIN: frzpower



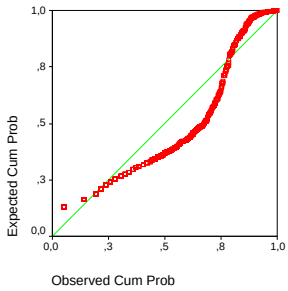
Weibull P-P Plot of DELTAHEU

LOGIN: GaëlleM



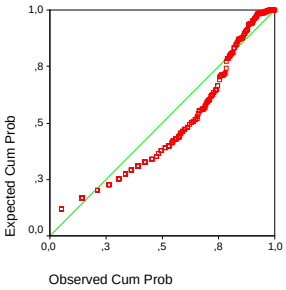
Weibull P-P Plot of DELTAHEU

LOGIN: gallouin



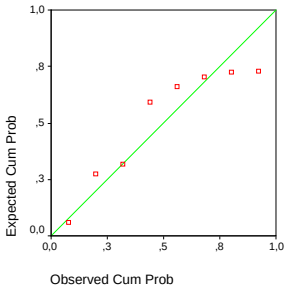
Weibull P-P Plot of DELTAHEU

LOGIN: Ginie31



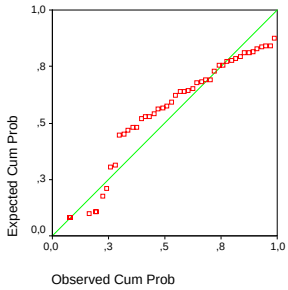
Weibull P-P Plot of DELTAHEU

LOGIN: gougay



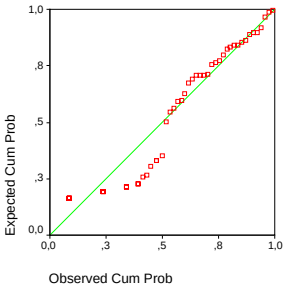
Weibull P-P Plot of DELTAHEU

LOGIN: Greensky



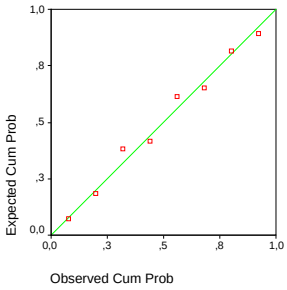
Weibull P-P Plot of DELTAHEU

LOGIN: gzing



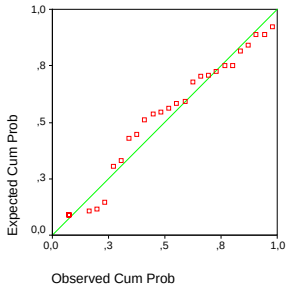
Weibull P-P Plot of DELTAHEU

LOGIN: hedi.dc



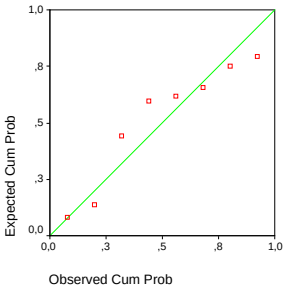
Weibull P-P Plot of DELTAHEU

LOGIN: hutch



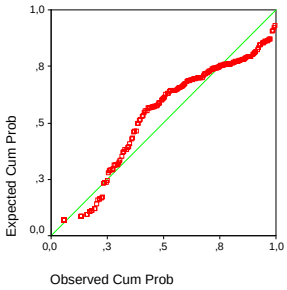
Weibull P-P Plot of DELTAHEU

LOGIN: jeanbask



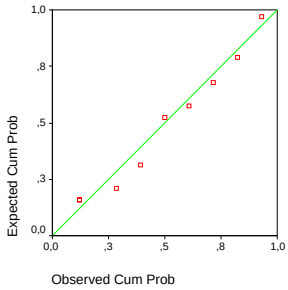
Weibull P-P Plot of DELTAHEU

LOGIN: jeannot3



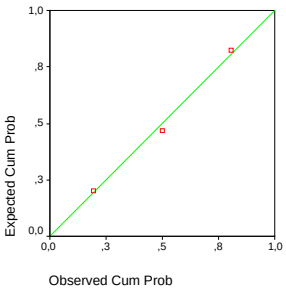
Weibull P-P Plot of DELTAHEU

LOGIN: joewj



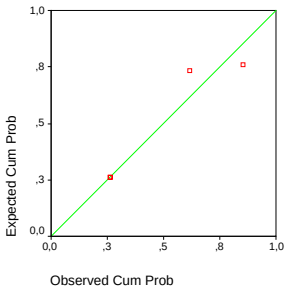
Weibull P-P Plot of DELTAHEU

LOGIN: juboubou



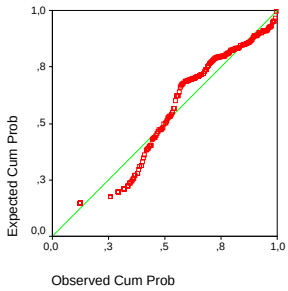
Weibull P-P Plot of DELTAHEU

LOGIN: juju75



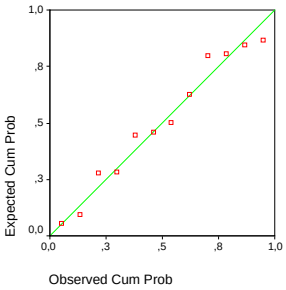
Weibull P-P Plot of DELTAHEU

LOGIN: kabotine



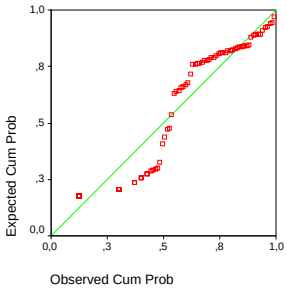
Weibull P-P Plot of DELTAHEU

LOGIN: Kathrin



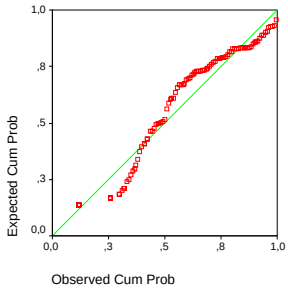
Weibull P-P Plot of DELTAHEU

LOGIN: Kifran



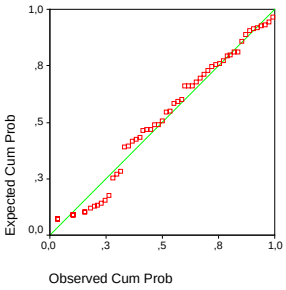
Weibull P-P Plot of DELTAHEU

LOGIN: lbarbier



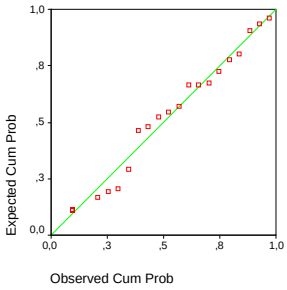
Weibull P-P Plot of DELTAHEU

LOGIN: levieux



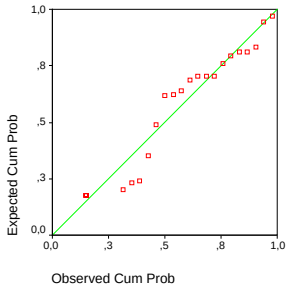
Weibull P-P Plot of DELTAHEU

LOGIN: LoïcT



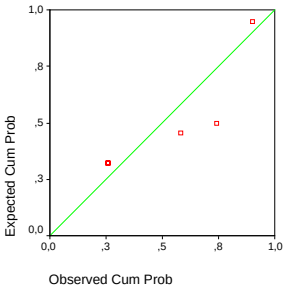
Weibull P-P Plot of DELTAHEU

LOGIN: Loquet



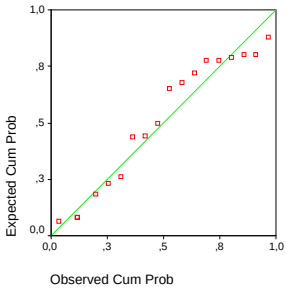
Weibull P-P Plot of DELTAHEU

LOGIN: luberdov



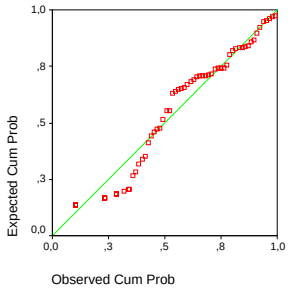
Weibull P-P Plot of DELTAHEU

LOGIN: ludovicb



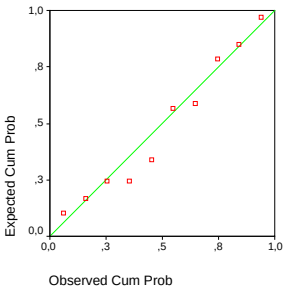
Weibull P-P Plot of DELTAHEU

LOGIN: m@hy



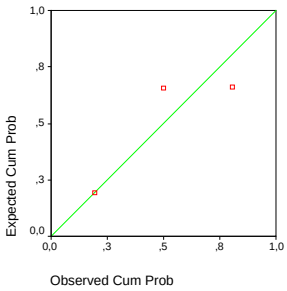
Weibull P-P Plot of DELTAHEU

LOGIN: manwe



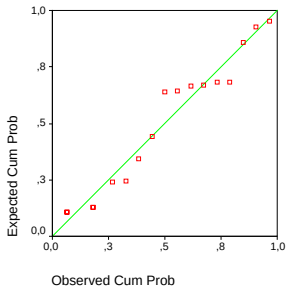
Weibull P-P Plot of DELTAHEU

LOGIN: mary_ce



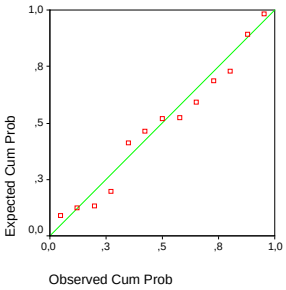
Weibull P-P Plot of DELTAHEU

LOGIN: Mayfair



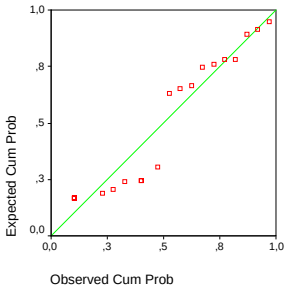
Weibull P-P Plot of DELTAHEU

LOGIN: mekiela



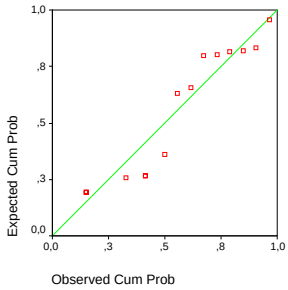
Weibull P-P Plot of DELTAHEU

LOGIN: Michel



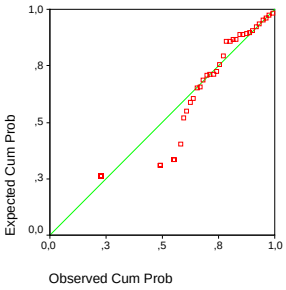
Weibull P-P Plot of DELTAHEU

LOGIN: micka74



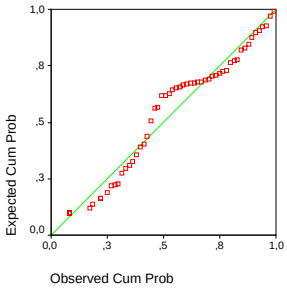
Weibull P-P Plot of DELTAHEU

LOGIN: mike_s



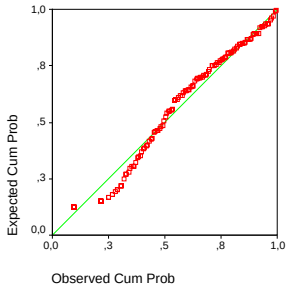
Weibull P-P Plot of DELTAHEU

LOGIN: minious



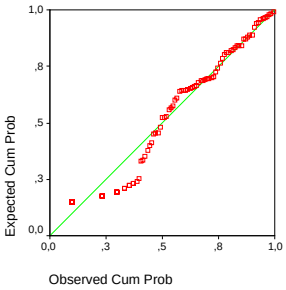
Weibull P-P Plot of DELTAHEU

LOGIN: mperrot



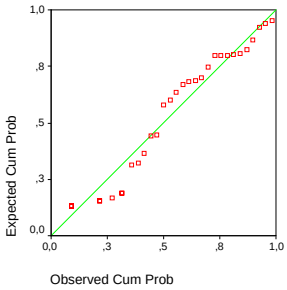
Weibull P-P Plot of DELTAHEU

LOGIN: Munier



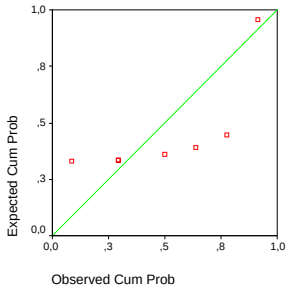
Weibull P-P Plot of DELTAHEU

LOGIN: Naemesis



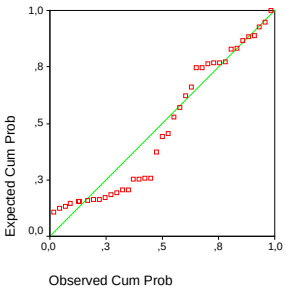
Weibull P-P Plot of DELTAHEU

LOGIN: nhai07



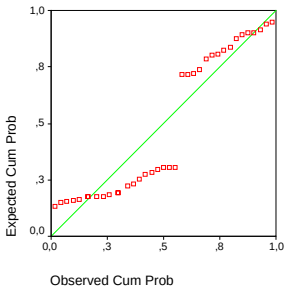
Weibull P-P Plot of DELTAHEU

LOGIN: nickus



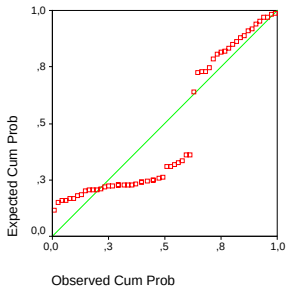
Weibull P-P Plot of DELTAHEU

LOGIN: noel_net



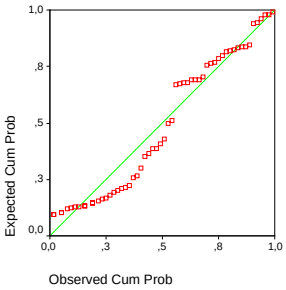
Weibull P-P Plot of DELTAHEU

LOGIN: philia



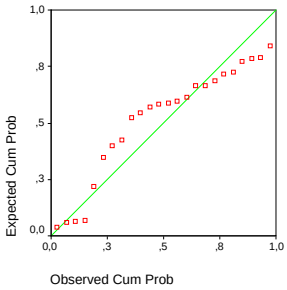
Weibull P-P Plot of DELTAHEU

LOGIN: popcorn



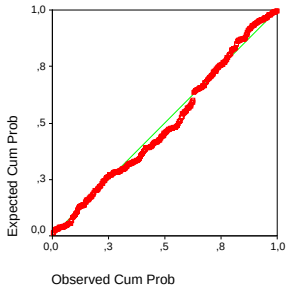
Weibull P-P Plot of DELTAHEU

LOGIN: prioton



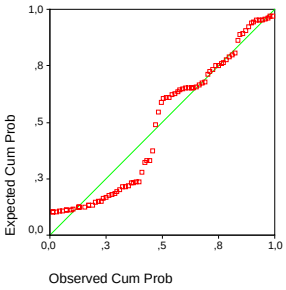
Weibull P-P Plot of DELTAHEU

LOGIN: prozaiik



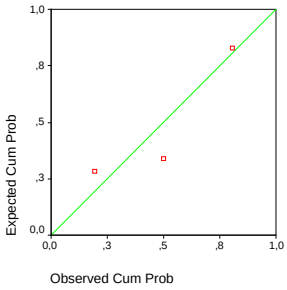
Weibull P-P Plot of DELTAHEU

LOGIN: quentin



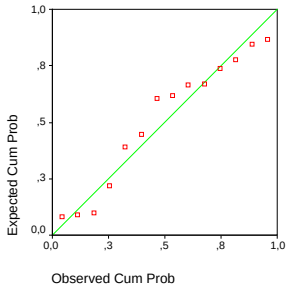
Weibull P-P Plot of DELTAHEU

LOGIN: RAPH



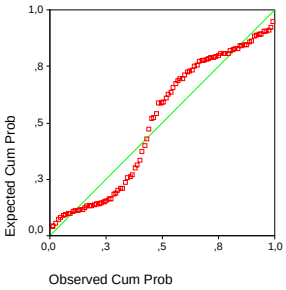
Weibull P-P Plot of DELTAHEU

LOGIN: relindis



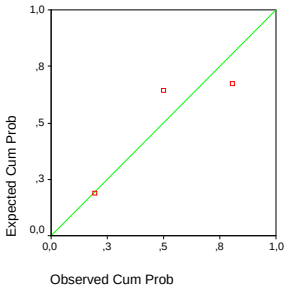
Weibull P-P Plot of DELTAHEU

LOGIN: rlaval



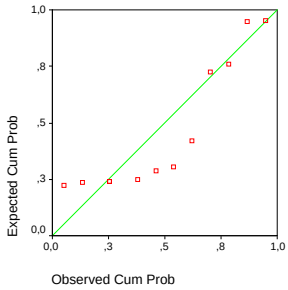
Weibull P-P Plot of DELTAHEU

LOGIN: sandor



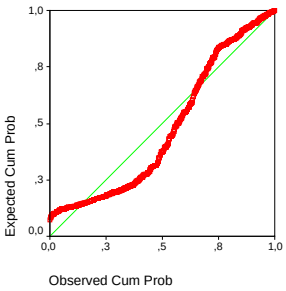
Weibull P-P Plot of DELTAHEU

LOGIN: sebulba9



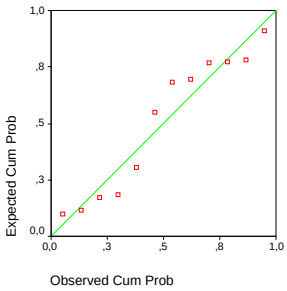
Weibull P-P Plot of DELTAHEU

LOGIN: semery



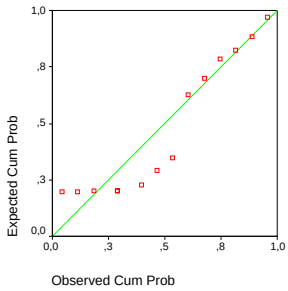
Weibull P-P Plot of DELTAHEU

LOGIN: sicart



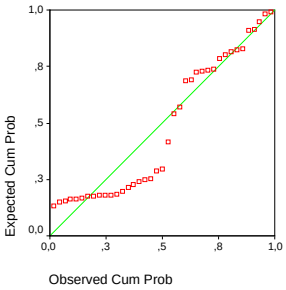
Weibull P-P Plot of DELTAHEU

LOGIN: sophynet



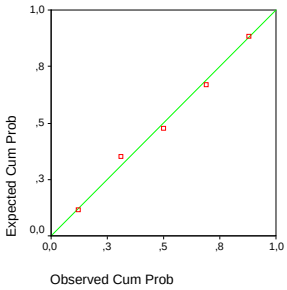
Weibull P-P Plot of DELTAHEU

LOGIN: Stèph



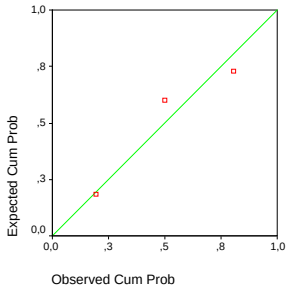
Weibull P-P Plot of DELTAHEU

LOGIN: stumpf



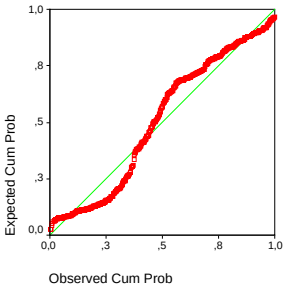
Weibull P-P Plot of DELTAHEU

LOGIN: Sylvie4



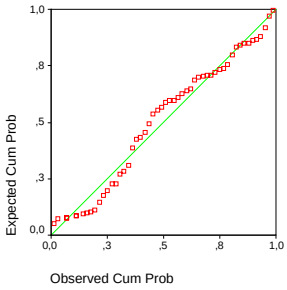
Weibull P-P Plot of DELTAHEU

LOGIN: Thibaut



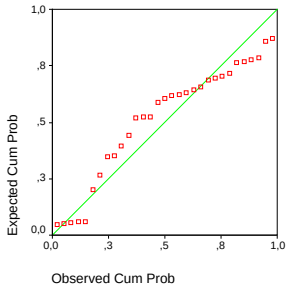
Weibull P-P Plot of DELTAHEU

LOGIN: Tonio2k



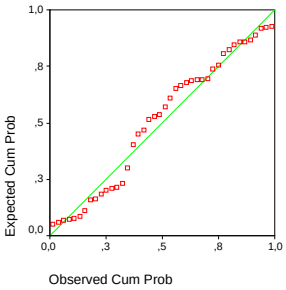
Weibull P-P Plot of DELTAHEU

LOGIN: toto2



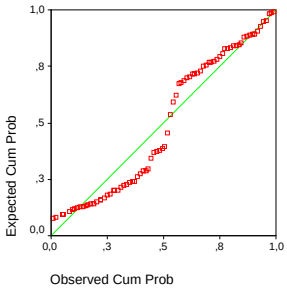
Weibull P-P Plot of DELTAHEU

LOGIN: tranb



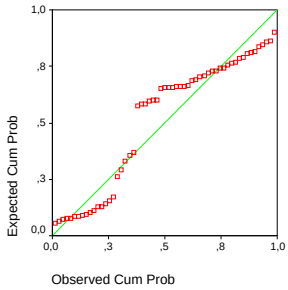
Weibull P-P Plot of DELTAHEU

LOGIN: vianney2



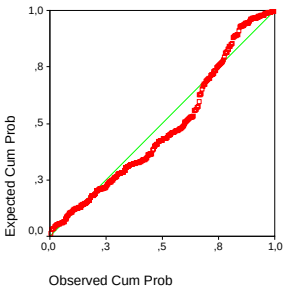
Weibull P-P Plot of DELTAHEU

LOGIN: Virginie



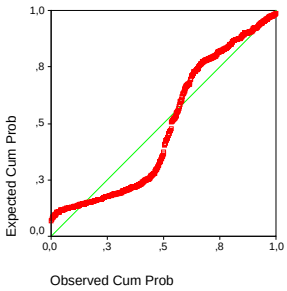
Weibull P-P Plot of DELTAHEU

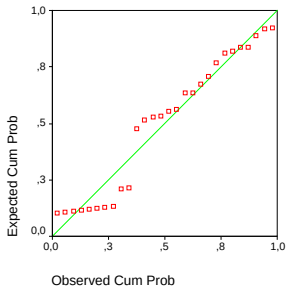
LOGIN: xzouf



Weibull P-P Plot of DELTAHEU

LOGIN: Yseulys



Weibull P-P Plot of DELTAHEU

LOGIN: Yves